A FRAMEWORK FOR VALIDATING REUSABLE BEHAVIORAL MODELS IN ENGINEERING DESIGN

A Thesis
Presented to
The Academic Faculty

by

Richard J. Malak Jr.

In Partial Fulfillment
Of the requirements for the Degree
Master of Science in Mechanical Engineering

Georgia Institute of Technology
March, 2005

A FRAMEWORK FOR VALIDATING REUSABLE BEHAVIORAL MODELS IN ENGINEERING DESIGN

Approved by:

Dr. Christiaan J. J. Paredis, Chair
School of Mechanical Engineering
Georgia Institute of Technology

Dr. Farrokh Mistree
School of Mechanical Engineering
Georgia Institute of Technology

Dr. Kwok-Leung Tsui
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Janet K. Allen
School of Mechanical Engineering
Georgia Institute of Technology

Date Approved: March 2005

ACKNOWLEDGEMENTS

No person is an island. Each of us is a complex product of our experiences and interactions with others. Furthermore, few achievements come without opportunities afforded to us by others. As such, it would be inappropriate for me to claim total credit for the completion of this thesis.

Above all others, I must thank my family for their years of support and encouragement. My mother has been particularly important to me. Looking back on my childhood, her devotion to my education is clear: Before I could read, she would read to me. Once I could read on my own, she would drive me regularly to the nearest public library—nearly an hour from our home. Throughout my college years, she has been a steady source of encouragement and has consistently filled the financial gaps left by my scholarships and loans. But perhaps most importantly, she instilled in me at an early age that I could do anything I wanted with my life—that I am not constrained by the paths taken by my ancestors and predecessors, but by the limits of my imagination and determination.

I must also acknowledge the roles of my younger siblings, my brother Daniel and my sister Amanda. Their support has been invaluable to me. They keep me grounded and remind me of what is important like no other people in my life. I value greatly my regular conversations with them. They are the best lifelines off of my island of academia.

To Donna Yocum, the person with whom I share my life, I offer the deepest thanks. It is hard to articulate the role she plays in my life or the impact that she has had on this work. Always there for me, she is the difference-maker in my life and the enabler of my career.

Academically speaking, I am most deeply indebted to my advisor, Dr. Chris Paredis. He is simultaneously open minded and critical—a combination of traits that is as valuable as it is rare. He has allowed me to explore new ideas and question existing ones, while consistently putting my own ideas to the test. His insights and criticisms have helped to shape both the ideas in this thesis and my broader understanding of engineering.

I also thank my reading committee: Dr. Farrokh Mistree, Dr. Janet Allen and Dr. Kwok-Leung Tsui. They challenged me with questions that ultimately lead this thesis to be more understandable and complete.

I owe gratitude to all the members of the Systems Realization Laboratory here at Georgia Tech. Various members of the SRL have provided me with valuable feedback as the ideas in this thesis have come together. Several members stand out as having been particularly helpful. Jason Augenbaugh, Steve Rekuc, Tarun Rathnam, Jay Ling, Morgan Bruns and Greg Mocko have all provided insightful comments on a regular basis. I thank them for their input.

Finally, I gratefully acknowledge the support of the G.W. Woodruff School of Mechanical Engineering and NASA Ames Research Center under cooperative agreement NNA04CK40A.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GLOSSARY

Adequacy Assessment A process by which one determines whether the accuracy of a behavioral model is adequate for a particular use.

Aleatory Uncertainty A type of uncertainty due to inherent variability.

Behavior What a system does; its performance. Depends on its form and the situation in which it is used.

Behavioral Attribute A subset of system behavior. Often relates to a an output of a behavioral model that is of interest in a decision problem.

Behavioral Model A representation of a system that maps attributes of form and use into a prediction about system behavior.

Behavioral Model Validation The process of determining whether a behavioral model is sufficiently accurate for user needs in the context of intended model scenarios.

Behavioral Prediction An output of a behavioral model.

Compatibility Assessment A process by which one determines whether the context of a behavioral model is a superset of the context of a particular problem (i.e., whether it is context-compatible).

Conceptual Model Validation The process of substantiating that the theories and assumptions underlying a conceptual model are correct and that the representation of the system is reasonable for the intended purpose of the model.

Context A formal specification of the circumstances under which a statement is has meaning.

Context Compatible A behavioral model is context compatible with a particular use situation if and only if the context of the behavioral model subsumes that of the use situation.

Data Validation The process of ensuring that the data necessary for modeling and simulation activities (model building, evaluation and testing, simulation, etc.) are adequate and correct.

Engineering Designer Someone who contributes to an engineering design process, possibly by being a model creator or model user.

Epistemic Uncertainty A form of uncertainty due to lack of knowledge.

Experiment A model and a set of inputs to that model that define a particular usage situation.

Form The physical characteristics of a system—i.e., what it is—including its geometry, topology and materials.

Function The expectations for a system—i.e., what it should do—expressed in an implementation independent fashion.

Inaccuracy An appraisal of the total uncertainty in a model or prediction.

Knowledge Relationships between different concepts or information. Can take the form of rules or equations, although not necessarily expressed explicitly (i.e., can be in the mind of an individual).

Model Something that someone can use to answer questions of interest about a particular system.

Model Creator Anyone who is involved in the development of a behavioral model for use in an engineering design process.

Model Reuse The act of applying a behavioral model to a situation other than the specific situation for which it was originally developed.

Model User Anyone who is involved in the (re)use of a behavioral model in an engineering design process.

Model Verification The process of assuring that a computerized model is a sufficiently accurate representation of the corresponding conceptual model.

Referent A source of knowledge that serves as "ground truth" for the purposes of model validation. Several knowledge sources can serve as a referent, including the first principles of a phenomenon, empirical data about the system and the opinions of a suitable domain expert.

Reuse See *model reuse*.

Simulation An experiment performed on a model. Typically involves the solution of a model using an algorithm implemented on a computer.

Simulation Validation The process of establishing that the results from a simulation study are sufficiently accurate within the intended set of situations. Includes model validation as a sub-process.

Validation-relevant Knowledge That which one must know in order to perform model validation. This includes knowledge about the limitations of a model, the system it represents and the objectives of the simulation study.

Validity Characterization A process by which one defines a validity description for a behavioral model.

Validity Description Formal and unambiguous statements about a behavioral model or prediction inaccuracy and the context over which one can trust the inaccuracy statement.

Verification See *model verification*.

NOMENCLATURE

C_i	Context for item labeled 'i'.
$\mathbf{F}(t)$	Net force vector.
m	(Constant) Particle mass.
$\mathbf{a}(t)$, $\dot{\mathbf{v}}$	Acceleration vector.
$m(t)$	Time-varying particle mass.
$\mathbf{v}(t)$	Velocity vector.
$\dot{m}$	Time-derivative of mass.
$\lVert\cdot\rVert_2$	Euclidean norm: $\lVert\mathbf{v}\rVert_2 = \left(\mathbf{v}^T\mathbf{v}\right)^{1/2}$ for a vector, $\mathbf{v}$.
$\beta_{\dot{m}}$	Bound on the rate-change of mass.
$\beta_{\mathbf{v}}$	Bound on the norm of the velocity.
$\mathbf{e}$	Vector error term.
β	Inaccuracy bound for a model.
$\sigma, \tilde{\sigma}$	Engineering stress and true stress, respectively.
$\varepsilon, \tilde{\varepsilon}$	Engineering strain and true strain, respectively.
E	Young's modulus.

A_0, A	Respectively, initial and final cross-sectional area of a material sample / beam.
L_0, L	Respectively, initial and final length of a material sample / beam.
α	Linear thermal expansion coefficient.
T_i, T_f	Initial and final temperature, respectively.
δ_{Hooke}	Scalar error term for Hooke's law
δ_{Beam}	Scalar error term for model of beam held in axial tension.
F	Applied loading force for beam.

SUMMARY

Designers commonly use computer-based modeling and simulation methods to predict artifact behavior. Such predictions are central to engineering decision making. As such, determining how well they correspond to actual artifact behavior is a problem of critical importance. A significant aspect of this problem is determining whether the model used to generate the behavioral predictions—i.e., the *behavioral model*—reflects the relevant physical phenomena. The process of doing this is referred to as *behavioral model validation*.

Prior works take an integrated approach to validation in which model creators and model users interact throughout the modeling and simulation process. Although effective for many problems, this type of approach is not appropriate for model reuse scenarios. Model validation requires knowledge about the model and its use. In model reuse scenarios, model creators and model users operate in independent processes with limited inter-process communication. The core challenge to behavioral model validation in this setting is that, in general, neither model creators nor model users possess the requisite knowledge to perform behavioral model validation.

Presented in this thesis is a conceptual framework for validating reusable behavioral models in model reuse scenarios. This framework solves the problem of creator-user separation by defining specific validation responsibilities for each and an interface by which they communicate. This interface consists of a formal description of

the model's limitations and the domain over which these limitations are known to be true. The framework is illustrated through basic engineering examples.

CHAPTER 1:

BEHAVIORAL MODEL REUSE AND VALIDATION

How can engineering designers establish trust in a behavioral model? Behavioral models are elemental to engineering design. They are the basis for all predictions about artifact behavior—i.e., all assessments of behavior other than those made by observing the artifact itself. Because designers use these predictions when making decisions, behavioral models have a direct impact on a design and when used inappropriately may lead to designs that fail to meet intended functionality. Accordingly, engineering designers must understand the limitations of the models they use. The process of establishing this trust is known as *model validation* (Schlesinger, et al. 1979).

Model validation is a knowledge-intensive endeavor and the organization of relevant knowledge influences how validation can be performed. Model validity depends on several factors, including the assumptions embodied in the model, the simulation experiment scenario and the intended use of the resulting predictions. Model validation is relatively straightforward when one person has a deep understanding of all the relevant knowledge. However, this is not the case in all engineering design situations. The focus of this thesis is on behavioral model validation for scenarios in which a behavioral model is *reused by someone other than its creator*. Such scenarios potentially result in a separation between those who have knowledge relevant to model validation and those who require that knowledge. One can understand how this happens by considering the interactions between model creators and model users.

In this thesis, *model creator* refers to one who develops a behavioral model for use in an engineering design process. A *model user* is one who applies a behavioral model to a specific task in a design process. When an engineering designer[1] is both creator and user of the same model, there is no separation of validation-relevant knowledge. However, the situation is more complicated when a model user is not its creator. Model creators know the details of a model and its limitations while model users know the details of an intended simulation scenario and use of the predictions. Furthermore, one cannot assume that a model user and model creator will be able to interact. For instance, the creator of a model may no longer work for the company that owns the model.

Chapter 2 contains an explanation of how existing approaches to model validation are insufficient to deal with the realities of behavioral model reuse. The core problem is that the conceptual framework underlying these approaches lacks appropriate concepts to describe reuse scenarios. There exists a need for a new framework in which there is a clear distinction between the roles and responsibilities of model creators and model users and an explicit treatment of validation-relevant knowledge. Engineering designers should be able to use the framework to conceptualize the problem of reusable behavioral models, but it should leave them free to develop specific methods and knowledge representations that suit their individual needs. This is in recognition of the complexity of behavioral model validation. A "one-size-fits-all" method for the validation of reusable behavioral models is unlikely to exist, but there is a framework of concepts and

[1] In this thesis, *engineering designer* refers to anyone who contributes to an engineering design process. This includes anyone who is a creator or user of behavioral models for a design process.

responsibilities that are common to all successful approaches. This thesis describes such a framework.

This chapter is devoted to describing and motivating the problem of validating reusable behavioral models and to explaining the structure of the remainder of this thesis. Section 1.1 is a discussion of behavioral models with the objective of identifying the types of interest in this thesis. Section 1.2 is an introduction to the problem of model validation and the notion of validation-relevant knowledge. It includes definitions of relevant terms and a brief discussion about the knowledge involved in model validation. Section 1.3 is a description of behavioral model reuse. It contains an explanation for why behavioral model reuse is important in engineering design, descriptions of different reuse scenarios and a discussion of the challenges that reuse poses behavioral model validation. Section 1.4 contains descriptions of the specific research questions to be addressed in this thesis, the corresponding hypotheses and a discussion about how the hypotheses are evaluated. Section 1.5 is an account of the structure of this document and includes a roadmap identifying the objectives of each chapter.

1.1 Behavioral Models

The focus of this thesis is on a particular type of model: symbolic behavioral models. The objective of this section is to describe what this means. The notion of a model as used in this thesis is defined in Section 1.1.1. Different types of model implementations are identified in Section 1.1.2. The meaning of 'behavioral model' is explained in Section 1.1.3.

1.1.1 Models

Although there exists no consensus definition for the term ‘model,’ most authors agree on two things: the function of a model is to serve as a surrogate for something else, and that, at some level of detail, models differ from the things that they represent. A sampling of definitions from the modeling and simulation literature is given in Table 1.1. The following definition is used in this thesis.

> Model: Something that someone can use to answer questions of interest about a particular system.

This definition is adapted from a similar one given by Minsky (Minsky 1965), which is stated in its original form in Table 1.1. The adopted definition does not restrict what can serve as a model. It stipulates only that a model serves as a surrogate for another system for the purposes of gathering information about it. This definition is chosen because, unlike several of those in Table 1.1, it accentuates the purpose of a model. Understanding models in terms of their purpose helps one to understand the role of models (behavioral or otherwise) in engineering design.

1.1.2 Symbolic Models

The definition for ‘model’ adopted in this thesis, as well as those given in Table 1.1, place no restrictions on *what* can serve as a model. Given a question that one wishes to answer, there may be several ways to realize a corresponding model. The arguments presented in this thesis focus on a particular class of behavioral models—referred to below as *symbolic models*. Hazelrigg describes three classes of things that serve as

Table 1.1: Some definitions for "model."

Definition	Reference
"A representation of a physical system or process intended to enhance our ability to understand, predict or control its behavior."	(AIAA 1998)
"A *model* is a representation or abstraction of something such as an entity, a system or an idea."	(Balci 2001)
"The facility of process of interest is usually called a *system*, and in order to study it scientifically we often have to make a set of assumptions about how it works. These assumptions, which usually take the form of mathematical or logical relationships, constitute a *model* that is used to try to gain some understanding of how the corresponding system behaves."	(Law, et al. 2000)
"To an observer B, an object A* is a model of an object A to the extent that B can use A* to answer questions that interest him about A."	(Minsky 1965)
"A physical, mathematical or otherwise logical representation of a system, entity, phenomenon or process."	(US DoD2003)
"The most common concept of a simulation model is that it is a set of instructions, rules, equations, or constraints for generating I/O behavior."	(Zeigler, et al. 2000)

models that are notable in the context of engineering design (Hazelrigg 1999): iconic, analog and symbolic.

Iconic models are physical representations of the real world. Examples include model ships, model bridges and model airplanes. They include both scaled (up or down) and non-scaled models. Non-scaled iconic models are more commonly referred to as *prototypes*. Iconic models are still used in engineering design, but have been replaced by computer-based approaches in many situations.

Analog models are those that represent one phenomenon using similarities in another phenomenon. The objective is to use a more convenient phenomenon to represent another. For instance, one can represent a second-order system using an electrical circuit comprised of resistors, capacitors and inductors. This is useful because it often is more expensive and time consuming to build the target system than it is to construct an equivalent RLC circuit. It also is common for engineers to make use of mental analogies during informal qualitative reasoning. For example, a mechanical engineer relies on his or her intuition of a spring-mass-damper system (a second-order system) in order to think about passive electronics (RLC circuits).

Symbolic models are those that in which one uses symbols to represent physical quantities and express the relationships between these quantities in mathematical forms. The majority of modern engineering models are symbolic models. Hazelrigg attributes this in part to the advent of digital computers that enable fast analysis of such models. It is also due to the breadth of questions that one can answer using them. For example, one can predict quantities using symbolic models that are not directly observable in the target system. Symbolic models are completely abstracted from the system of interest. As

Hazelrigg notes, "there is no correspondence between a symbolic model and the reality that it is intended to represent other than that which is in the head of the modeler." This allows the greatest flexibility in developing and using symbolic models, but also has the drawback making it easy to develop incorrect models or to misuse them. Symbolic models are the focus of this thesis.

1.1.3 Behavioral Models

To describe a model as a 'symbolic model' conveys what is used as a representation. To describe a model as a 'behavioral model' conveys what is being represented. Thus, behavioral models do not necessarily differ structurally from other types of models.

Behavior is one of three concepts commonly used to describe knowledge and information in an engineering design process. The other two are *form* and *function*. The three concepts are used in this thesis as follows:

> Form: The physical characteristics of a system—i.e., what it *is*—including its geometry, topology and materials (Clayton, et al. 1999, Shooter, et al. 2000) (Gero refers to this same concept as *structure* (Gero 1990)).
>
> Function: The expectations for a system—i.e., what it *should do*—expressed in a implementation-independent manner (Gero 1990, Clayton, et al. 1999, Szykman, et al. 2001).
>
> Behavior: The performance of a system—i.e., what it *does*—in the context of its intended use (Gero 1990, Clayton, et al. 1999, Shooter, et al. 2000).

One way to think of form is as the knowledge required to manufacture and build a system. For a basic mechanical system, this might consist of dimensioned drawings of the parts to be machined and instructions for assembling the parts. Function often is defined in abstract, implementation-independent terms such as energy, mass and signal flows (e.g., the function structures of (Pahl, et al. 1996)). Behavior depends on several factors, including the physical characteristics of a system (i.e., its form) and the circumstances in which it is used. It consists of quantitative measures of system performance under specific conditions (Clayton, et al. 1999) that can be derived from physical principles (Shooter, et al. 2000). Behavior has to do with a particular system in a particular situation.

An engineering designer's task is to identify the most preferred design (i.e., form) that meets the specified function. The behavior of a design alternative determines whether it meets the specified function and the degree to which it is preferred. Engineering designers use behavioral models to predict the behavior of design alternatives. A behavioral model maps the form of an alternative and a usage scenario into a prediction about a particular behavioral attribute. Generally, one system has several behavioral models associated with it. Each of these models predicts different behavioral attributes under different circumstances. For example, the behavioral models associated with an airplane design might include one that predicts aerodynamic drag, one that predicts fuel consumption at a particular airspeed and several that predict aircraft stability through different maneuvers.

1.2 The Validation Problem and Validation-Relevant Knowledge

The purpose of this section is to describe the problem of behavioral model validation as considered in this thesis, but absent the issue of model reuse. An exploration of model reuse and its impact on model validation is presented in the next section. By presenting these ideas in sequence, it is easier to convey the implications of model reuse on behavioral model validation.

This section begins in Section 1.2.1 with a conceptual example intended to relate model validation to engineering design and illustrate some of the relevant concepts. The concepts of validity, model validation and validation-relevant knowledge are defined more precisely in Section 1.2.2. Section 1.2.3 contains remarks about the distinction between model validation and simulation validation.

1.2.1 A Conceptual Example

Behavioral model validation arises in many aspects of engineering design. To help visualize the model validation problem, it is informative to consider a design-inspired example scenario:

> A customer is seeking a design for a new space probe with particular functionality. The customer prefers to maximize its lifetime while meeting the functionality requirements. Designers have identified a set of alternatives that meet the desired functionality, but that may have differing lifetimes. The problem the designers face is to determine which alternative is most preferred.

Because it is not practical to build and evaluate each alternative, the designers must use a behavioral model to predict alternative lifetimes. An appropriate behavioral model maps characteristics of the alternative (e.g., fuel capacity, mass, structural characteristics, power requirements, battery characteristics, etc.) and characteristics of its use (e.g., planned maneuvers, radiation exposure levels, incident light intensity, etc.) into a prediction of its lifetime. An appropriate behavioral model also yields lifetime predictions that allow designers to differentiate between the alternatives.

The problem of model validation is to determine whether a particular behavioral model is appropriate for a particular use. With respect to the space probe example, there are two questions that must be answered in order to determine model validity:

1. Can the model yield lifetime predictions that correspond to the intended physical situation?
2. Can designers make a rational decision using the lifetime predictions yielded by the model?

The first question relates to whether one can use the model to generate predictions with the correct meaning. In this example, designers must use a behavioral model that is consistent with the customer's assumptions about the scenarios in which the space probe is used. For example, the significance of factors such as solar radiation, aerodynamic drag, temperature changes and magnetic effects depends on the particular mission plan. A mission that involves orbiting a planet with a significant atmosphere requires a fuel consumption model (which is a factor in probe lifetime) that accounts for orbit maintenance maneuvers necessitated by aerodynamic drag. In contrast, aerodynamics are not a significant concern for a deep-space probe. Designers must choose a model that corresponds to the mission as defined by the customer.

Assuming the meanings of attribute predictions are consistent with intentions, the usefulness of predictions comes in to play. This is the topic of the second question. In this example, useful predictions are those with which designers can rationally identify the most preferred alternative. A model that yields inaccurate lifetime predictions may preclude designers from discriminating between the alternatives. To chose an alternative in this case would be irrational. However, high accuracy is not a necessary condition for usefulness. It may be possible for designers to discriminate between alternatives using an inaccurate model if the true differences in lifetime are large relative to the modeling inaccuracies. In general, the required level of accuracy depends on the particulars of a given problem. Designers might require low or moderate accuracy when comparing different design concepts, but high accuracy when comparing minor modifications to the same concept. Alternately, both alternatives may perform so well that the expense of differentiating between them is greater than any potential benefit of identifying the better one. Designers must choose a model that is sufficiently accurate for their needs.

1.2.2 Validity, Model Validation and Validation-Relevant Knowledge

The Oxford English Dictionary contains the following entry (1989):

> Valid:
> 1 : Good or adequate in law; possessing legal authority or force; legally binding or efficacious
> 2 : Of arguments, proofs, assertions, etc.: Well founded and fully applicable to the particular matter or circumstances; sound and to the point; against which no objection can fairly be brought
> 3 : Of things: Strong, powerful. (Archaic)
> 4 : Of persons: Sound or robust in body; possessed of health and strength. Also said of health.

The second definition is relevant to this thesis. According to it, a valid model is one that indisputably is suited for the needs of its user. Although absolute indisputability is a practical impossibility (see Section 2.1), it is possible to establish a framework within which all stakeholders can arrive at a consensus about model suitability. Thus, a *valid* *model is one that all participants agree is suitable for its intended use*. Presumably, they will agree that this is the case only when presented with evidence to that effect. Behavioral model validation is the process of gathering and analyzing such evidence. The following definition is adopted in this thesis:

> Behavioral Model Validation: The process of determining whether a behavioral model is sufficiently accurate for user needs in the context of intended model scenarios.

For convenience, the terms 'model validation' and 'behavioral model validation' are used interchangeably throughout the thesis.

In order to better understand this definition, it is helpful to consider the key statements individually:

- Process of determining. This means that behavioral model validation is an ongoing process. Behavioral model validation involves evidence gathering as well as interpretation and making determinations.
- Sufficiently accurate for user needs. This phrase relates to how a user applies predictions generated using the behavioral model. Users may require different levels of accuracy for different problems. One judges model validity relative to particular user needs.
- In the context of intended model scenarios. This phrase reflects the set of physical scenarios, or *context*, the user intends to represent with the behavioral model. They are analogous to a set of experiments that one would perform on a physical model. Model accuracy can vary from scenario to scenario. One judges model validity relative to particular model scenarios.

Thus, behavioral model validation is a process that involves properties of a model and its use. A model is valid relative to a particular system, user needs and context. To do perform model validation, one requires knowledge about:

- The System. One must know the contents of a system and the details of its use. This dictates the intended model scenarios and which phenomena are of significance.
- The Model. One must know what assumptions are embodied in a model and their impact on its accuracy. The assumptions must be acceptable for the system of interest and the accuracy must be sufficient for the study objectives.
- The Study Objectives. One must know the simulation study requirements and objectives. This dictates the required prediction accuracy.

Effective use of this *validation-relevant knowledge* is a key to performing behavioral model validation in engineering design.

Unless otherwise noted, “behavioral model validation” and “model validation” are synonyms in this thesis. Several definitions for “model validation” are presented in Table 1.2 Specific phrasings vary, but most are substantially equivalent to the definition adopted here. For example, some authors use the phrase “domain of applicability,” which refers to the set of conditions under which a model is evaluated and thus relates to context. Also, some authors refer to “study objectives” or “model’s intended purpose,” both of which relate to user needs. The definitions that differ from the one adopted in this thesis do so by being less explicit or less general. For instance, the United States Department of Defense (US DoD) includes in their definition the gauging of model accuracy but do not include the determination of accuracy sufficiency. This is because they include sufficiency determination in another process (accreditation) (US DoD2003).

Table 1.2: Some definitions for "model validation."

Definition	**Reference**
"The process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model."	(AIAA 1998)
"Model validation is substantiating that the model, within its domain of applicability, behaves with satisfactory accuracy consistent with the study objectives."	(Balci 1997)
Validation "is the process of determining whether the model, (either conceptual or simulation), is an 'adequate' representation of the system."	(Birta, et al. 1996)
"Validation is the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model."	(Caughlin 1995)
"The goal in validating a simulation code is to determine the degree to which the output of the code agrees with the actual behavior of a physical system in a specified situation. Because the criterion is real-world behavior, validation must involve comparison of the simulation code's output to experimental results."	(Hanson, et al. 2001)
"Validation is the *process* of determining whether a simulation model (as opposed to a computer program) is an accurate representation of the system, *for the particular objectives of the study*." (Emphasis in original.)	(Law, et al. 2000)

Table 1.2 (continued)

"Validation involves substantiating that the model, within its domain of applicability, behaves with satisfactory accuracy consistent with the objectives governing its use. It relates to the comparison of model behavior with system behavior."	(Page, et al. 1997)
Termed Operational Validation: "Determining that the model's output behavior has sufficient accuracy for the model's intended purpose over the domain of the model's intended applicability."	(Sargent 1985)
"Substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model"	(Schlesinger, et al. 1979)
"The process of determining the degree to which a model and its associated data are an accurate representation of the real world from the perspective of the intended uses of the model."	(US DoD2003)
Use definition given by (Schlesinger, et al. 1979) or (Sargent 1987).	(Knepell, et al. 1993, Meckesheimer, et al. 2001, Chandrashekar, et al. 2002)

1.2.3 Model Validation versus Simulation Validation

From the perspective of an engineering designer, behavioral models exist to generate behavioral predictions. This is because the predictions are what engineering designers ultimately use to make decisions. For instance, an engineer makes a decision not because "$F = ma$," but because F evaluates to a particular value given m and a. Models are necessary to generate a prediction, but they are not sufficient. Likewise, model validation is necessary, but not sufficient to ensure the validity of the resulting prediction. Highly inaccurate inputs can lead to an insufficiently accurate prediction even if the behavioral model has sufficient accuracy. Thus, a distinction is made in this thesis between *model validation* and *simulation validation*.

Simulation validation is the process of determining whether a simulation yields predictions that are sufficiently accurate for user needs over the entire problem context. One can think of this as ensuring *prediction validity*. This sometimes is referred to as credibility assessment of simulation results or, simply, credibility assessment (Balci 1987, Knepell, et al. 1993, Balci 1997). It is a holistic concept that depends on the properties of all elements of a simulation. This includes a model, its inputs and its parameters.

The distinction between simulation validation and model validation is that the former establishes *actual* validity of predictions whereas the latter establishes the *potential* for prediction validity. Moreover, a model that is valid in a particular situation *can* produce a prediction that is valid in that situation provided it is part of an appropriate simulation. Thus, model validation is a necessary but insufficient *sub-process* of simulation validation.

The focus in this thesis is on model validation. It is a necessary precursor to making valid predictions about the behavior of a system. However, many of the ideas apply to the broader issue of simulation validation. Specific aspects of this link are discussed in Chapter 3.

1.3 Behavioral Model Reuse

In Section 1.2, model validation is described as a situated problem that requires specific knowledge about the system, the model and the simulation study objectives. Model validation cannot proceed if any of this validation-relevant knowledge is missing. However, this is precisely the problem encountered in the more general instances of model reuse. This section is an exploration of behavioral model reuse and its implications of model validation efforts. Section 1.3.1 is an exploration of the separation of validation-relevant knowledge from where it is required. This is accomplished through descriptions of different behavioral model reuse scenarios. These scenarios imply certain requirements that a model validation scheme must meet in order to support behavioral model reuse. Section 1.3.2 is a description of these requirements.

1.3.1 Reuse Scenarios

In order to validate a model, one must bring together various sources of relevant knowledge. The challenge with validating reusable models is that this knowledge can be far removed from where it is needed. To better understand this, it is useful to consider potential model reuse scenarios.

A model is reused by its creator

This is the simplest scenario for behavioral model reuse. A designer creates a behavioral model while working on one project and then reuses it in a similar situation on another project. Thus, this designer plays the roles of both a model creator and a model user. The designer is aware of the assumptions embodied in the model and understands the particulars of the model application and therefore can reach informed conclusions about model validity for the new situation. For instance, a designer might create a dynamics model under the assumption that a particular body is rigid. For the original application of the model, the body in question may have been very rigid. However, it may be that the body targeted in the new situation has a varying geometry. The designer is aware of the rigidity assumption in the model and is able to gauge the impact of violating it to some degree. Also, the designer is aware of the desired prediction accuracy and therefore can determine whether the model will yield sufficiently accurate predictions for the new body in question. In this reuse scenario, behavioral model validation is no more complicated than it is for the original use of the model.

A model is reused by someone other than its creator

In this scenario, a designer (the model user) seeks to reuse a model that was created by another designer (the model creator). The model user is responsible for validating the model with respect to his or her application. However, validation can be more difficult than in the preceding scenario. The reason for this is that there can exist a gap in knowledge between the model user and the model creator. Model users are familiar with the details of their applications, but not necessarily with the particular assumptions and limitations embodied in a particular behavioral model. This knowledge is essential to validation. Model creators understand the assumptions and limitations embodied in their models but not the particulars of all (or even most) potential uses of their models. Thus, neither model users nor model creators generally possess all the knowledge required to perform model validation.

The implication of this scenario is that, in general, model creators must exchange more than just a behavioral model with model users. Model creators must convey to model users any knowledge that is relevant to the model validation process. This additional exchange is depicted in Figure 1.1.

The details of exchanging validation-relevant knowledge from model creators to model users are complicated. In the simplest of situations, the creator and user work together closely. This allows them to exchange knowledge on an ad-hoc basis. However, close user-creator interaction is not possible in general. Interaction might be inconvenient because the model creator and model user do not work in the same place or because the model creator does not have time to help users. In more extreme situations, the model creator may not be available at all. This can occur for example if the creator leaves the company. Behavioral models can be stored indefinitely in a design repository (Szykman, et al. 1998, Szykman, et al. 2000) or a behavioral knowledge repository (Mocko, et al. 2004), which enables designers to access models without even communicating its creator. This increases the likelihood that a model user will have to perform model validation without working closely with the model creator. Ad-hoc strategies for the exchange of validation-relevant knowledge are not generally viable.

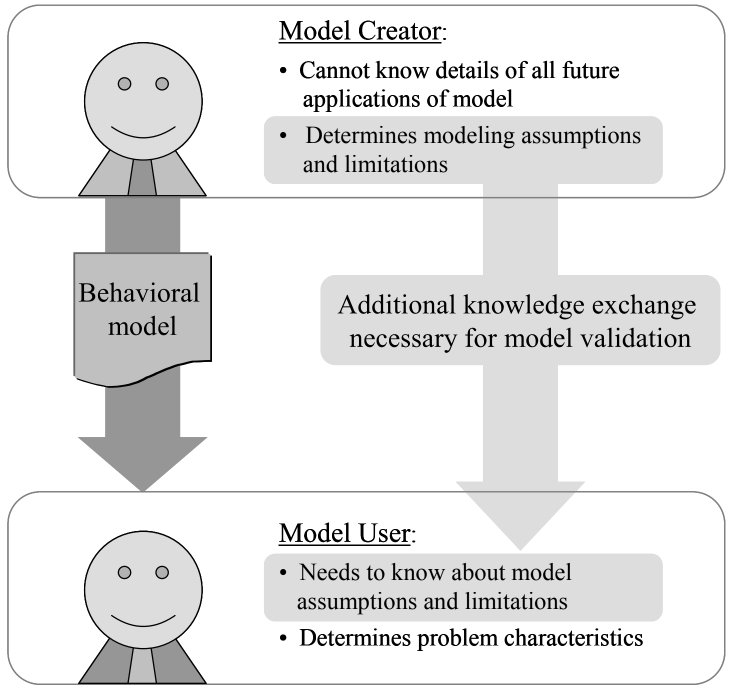

Figure 1.1: Knowledge exchanges for individual model reuse.

Compositional Modeling

In a compositional modeling scenario, a model user seeks to combine several models into a system model. Compositional modeling and model reuse are closely linked. By composing system models from simpler components, it is possible to enhance reuse and therefore model large systems in a cost-effective manner. Another benefit is that an individual model user can compose a model that includes phenomena from several disciplines by including the appropriate component models. An individual model creator typically is unable to develop highly complex multidisciplinary models.

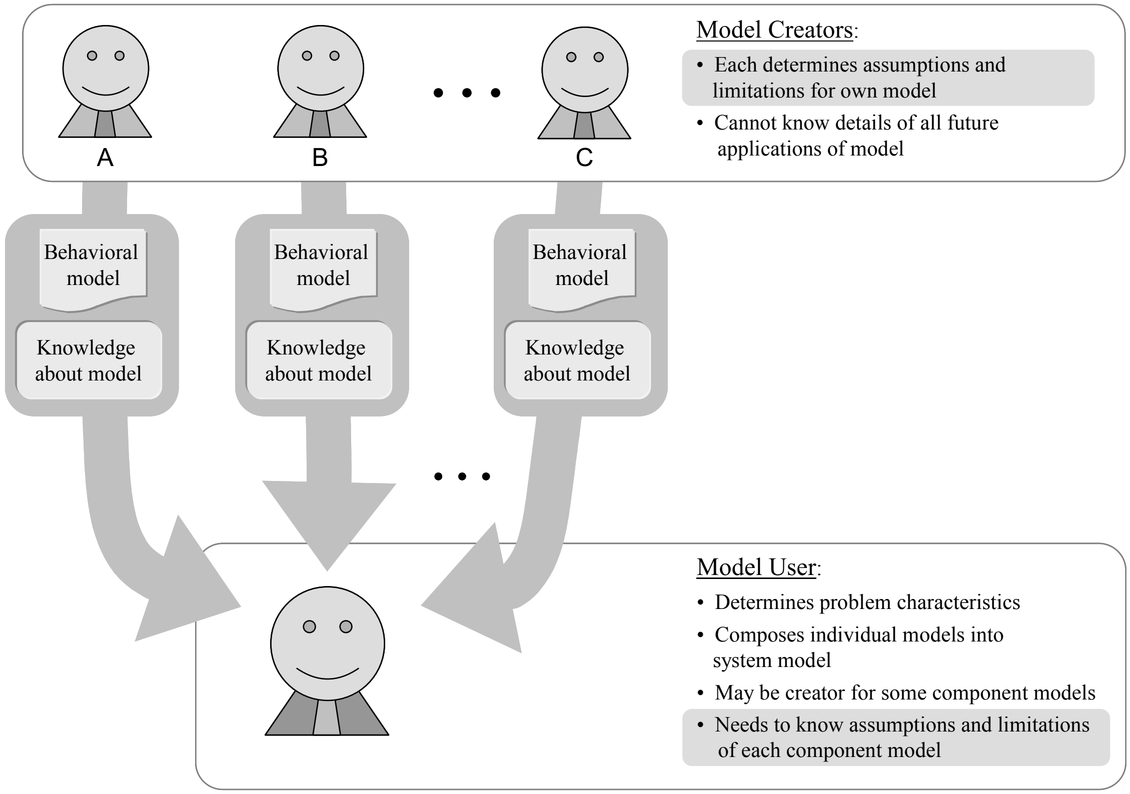

Figure 1.2 Knowledge exchanges for compositional modeling.

Compositional modeling presents many challenges for validation. Essentially, it is a scaled up version of the preceding scenario. The difference is that the model user utilizes models from many model creators. This is depicted in Figure 1.2. In this scenario, a model user may have created some of the models that go into the system model, but it is assumed that the user has not created all of them.

Validation of a composed model requires a superset of the knowledge required to validate each of its constituent models. Unlike the single-model scenario, the knowledge separation problem in compositional modeling is amplified by virtue of involving multiple component models. This results in two problems. First, a model user who uses a large number of models requires a large amount of knowledge about those models in order to perform model validation. This further restricts the modes of knowledge exchange. Where the single-model scenario precludes only ad-hoc knowledge exchange, compositional modeling requires efficiency in knowledge exchange. For example, traditional documentation (e.g., written language accompanied by some figures) is not a good way to exchange knowledge about the models in cases of compositional modeling. Even if each model is accompanied by a modest amount of documentation, the total can quickly grow out of hand for system models with moderate numbers of components. The second problem is that all of the models in a system model must be compatible with one another. Put differently, the component models must not contain assumptions that violate those of others. The most general way to ensure that a group of component models is not in conflict is to compare them on a pair-wise basis. However, this approach scales super-linearly with the number of component models and would be impractical to perform manually for any but the simplest of models. This suggests that a formal mathematical representation of modeling assumptions that can be used by an automated comparison routine would be preferred to a documentation-based knowledge representation.

It should be noted that one can think of the model user in this scenario also as a model creator in another scenario, since he or she composes individual models into a system model. A model user in the present scenario might publish a completed system model to a repository where it can be accessed by other model users. In this case, it would be beneficial if the model user were able to formulate efficiently the net limitations and assumptions for the system model based on knowledge about the component models. This factor also speaks to a formal knowledge representation.

Inter-organizational Model Exchange

In another scenario, behavioral models can be shared across corporate or organizational boundaries. This mode of reuse could be important in the defense industry where government agencies often perform situational analyses using models of their assets and personnel (e.g., war gaming, logistical analyses, etc.). Rather than modeling an asset (e.g., a tank, ship, building, etc.) themselves, the government agency can require a behavioral model as a deliverable from the contractor who designs and constructs the asset. This would be an efficient reuse of resources, since often the government requests a model to be developed as part of the design process.

In the abstract, this scenario is similar to the scenario in which a single user utilizes a model from a single creator. However, practical issues differentiate the two scenarios. In this scenario, a model is being reused by a different organization than the one that created it. This results in a separation of validation-relevant knowledge on a very large scale. Model creators and model users are in different organizations, which impedes (and possibly prevents) interaction, raises questions about trust and creates room for misinterpretations of the exchanged knowledge. These issues exist in any reuse scenario in which the model user is not its creator. However, inter-organizational exchange of behavioral models and knowledge about the models accentuates the existing issues.

Reuse of Predictions

In this scenario, a designer reuses a prediction for a new problem. Reuse of a prediction is analogous to reuse of a behavioral model. Like models, the validity of

Table 1.3: Requirements for schemes for validating reusable behavioral models.

No.	Requirement
1	The time spent performing validation activities at the point of model use must be made small.
2	Validation-relevant knowledge must be represented explicitly and associated with behavioral models.
3	Validation-relevant knowledge must be described in terms of concepts that have well-defined semantics that are independent of any particular person, group or project.
4	Validation-relevant knowledge must be expressed in a mathematically formal manner.

predictions is an important consideration. Because one uses models to compute them, predictions have associated assumptions and limitations. Designers must be able to formulate knowledge about a prediction in the same way they do for behavioral models.

1.3.2 Requirements for Behavioral Model Reuse

The reuse scenarios of the previous subsection point to several requirements that a model validation scheme must have in order to support efficient and effective behavioral model reuse. The requirements are summarized in Table 1.3 and explained below. They are a consideration when evaluating the hypotheses proposed in Section 1.4.1.

Time minimization at point of use

For behavioral model reuse to be worthwhile, it must be possible to reuse a model in much less time than it would take to develop a new one. Ignoring model validation, reuse is faster because it eliminates model development and leaves only simulation activities.

However, the difference between reuse and new development can be small if validation requires significant investments of time at each model use. Therefore, it is important to minimize the time required for at-use validation activities.

This requirement implies that all validation activities that are specific to the model (as opposed to a particular application of the model) should be performed up-front (i.e., when the model is developed rather than when it is used). Another implication is that one should avoid a validation scheme that requires significant creativity and insight at the point of model use. Such activities are time consuming and less predictable as compared to computational activities and should be conducted up-front whenever possible.

Association of explicit knowledge with models

Advances in information technology have made it easier for designers to access models created by others. This is good from an ease-of-reuse perspective, but complicates the model validation picture. What if the model creator is unavailable to assist in validating the model for its new use? It is not reasonable to expect a model user to reverse engineer a model in order to deduce its underlying assumptions and limitations. At best, such an effort is time consuming. At worst, it is prone to mistakes and could take longer than it would to develop a new model from scratch.

Model creators can associate with their models explicit representations of modeling assumptions and limitations. This frees model users to validate the use of a model without consulting its creator, thereby allowing model reuse to be robust to changes in creator accessibility.

Well-defined semantics for validation-relevant knowledge

Explicitly representing validation-relevant knowledge is just a first step. In order to be useful, the knowledge representation must be understandable by designers other than the model creator. This means that the concepts used to express the knowledge must have well-defined meanings. Furthermore, these meanings must be defined in terms that potential model users will understand. Defining model assumptions in terms of project-specific variables may be useful for the initial application of a model, but it is a hindrance to future reuse on other projects.

In addition to being expressed in terms of well-defined concepts, the interpretation of the validation-relevant knowledge in terms of its role in model validation must be clear to all participants. One can achieve this by appropriate definition of a model validation framework in which categories of knowledge are defined along with their roles in model validation.

Mathematical formality of knowledge representation

It is essential that validation-relevant knowledge be represented in an unambiguous fashion. Well-defined semantics solves this problem only in part by providing the concepts in the representation with an interpretation in the problem domain. Mathematical formality ensures that the relationships between the concepts are unambiguous and, therefore, computable and more easily enforced. This is the difference between saying "$mass_{structure}$ must be less than $mass_{max}$" and "$mass_{structure} < mass_{max}$." Assuming the meanings of the terms are well-defined, then both statements are semantically equivalent. However, the second form is combined more easily with other statements and can be evaluated computationally.

One should not confuse this requirement with *formal validation*, which is a way of using formal proofs of correctness to support model validation (Balci 1995).

1.4 Validation of Reusable Behavioral Models

The problem of behavioral model validation and the challenges associated with behavioral model reuse are described in the previous sections. Behavioral model validation requires specific validation-relevant knowledge in order to proceed effectively. Behavioral model reuse leads to a gap between where this knowledge exists and where it is needed. This section contains a description of how this problem is addressed in this thesis. Section 1.4.1 is an account of the research questions addressed and the corresponding hypotheses. Section 1.4.2 is a description of the approach adopted for evaluating the hypotheses.

1.4.1 Research Questions and Hypotheses

Recent advances in information technology and knowledge management promise to make behavioral model reuse easier and more widely available. Technologies such as design repositories (Balci 1995, Szykman, et al. 1998, Szykman, et al. 2000) and behavioral model repositories (Mocko, et al. 2004) allow users to search for and retrieve behavioral models without interacting with their creators. Although these technologies make it easier for designers to reuse behavioral models, they do nothing to help designers decide whether they *should* reuse a particular behavioral model in a particular situation. What is more, they expose a structural problem with behavioral model reuse: validation-relevant knowledge is divided between model creators and model users, two groups that may have no means of exchanging this knowledge directly. If engineering designers are to reuse behavioral models efficiently and correctly, the validation first must be overcome. This motivates the primary research question addressed in this thesis:

> ***Primary Research Question:*** *How can engineering designers perform behavioral model validation in a way that supports model reuse?*

To make this broad question more tractable, it is helpful to decompose it into sub-questions. The requirements described in Section 1.3.2 call for the efficient and effective formalization and use of validation-relevant knowledge. The sub-questions addressed in this thesis break down according to the two steps of *knowledge formalization* and *knowledge use*.

The first research question, Q1, deals with the problem of knowledge formalization. It reflects the requirements stated in Table 1.3. The question and the corresponding hypothesis, H1, follow:

> ***Q1:*** *How can model creators convey validation-relevant knowledge in a way that is independent of any person, group or project?*
>
> ***H1:*** *Model creators can develop mathematical descriptions of their creations—called validity descriptions—that provide assertions about the accuracy a user can expect and the context over which the assertions hold true.*

The notion of a validity description is a central concept in this thesis. It is a mechanism for the formal communication of validation-relevant knowledge about a model that is independent of any particular person, group or project. The process of developing a validity description is called *validity characterization*.

Although one could convey the same validation-relevant knowledge through systematic documentation, formal mathematical descriptions have several advantages. The most significant advantage is that they are unambiguous, which is important in cases where model creators are unavailable to provide clarification. Another advantage is that they potentially lead to faster model use. This is because they are amenable to automated processing.

The second research question, Q2, addresses the issue of knowledge use. It and the corresponding hypothesis, H2, follow:

> ***Q2:*** *How can model users apply validity descriptions to validate the application of a behavioral model to a particular problem?*
>
> ***H2:*** *Model users can perform a two-step assessment process in which they:*
>
> *(1) determine whether the context stated in the validity description is compatible with the problem and, if it does,*
>
> *(2) determine whether the accuracy stated in the validity description is sufficient for the needs of the problem.*

The first step in H2 is a check to see whether a model and a problem are compatible in the sense that the scenarios represented by the model correspond to those intended by the user. Moreover, one compares the context associated with a model to that of the simulation problem. This check is called *compatibility assessment* and is a necessary condition for model validity. A model that fails compatibility assessment is not valid for the problem being considered. The second step in H2 is a check to see whether a model is adequate for the needs of the problem in the sense that can yield predictions with sufficient accuracy as determined by the problem. This check is called *adequacy assessment* and also is a necessary condition for model validity. A model that fails adequacy assessment is not valid for the problem being considered. A model that passes both compatibility and adequacy assessment is valid for the problem being considered.

Together, H1 and H2 form a conceptual framework for the validation of reusable behavioral models. They consist of concepts, tasks, decisions and relationships that are necessary to achieve the objective of validating reusable behavioral models. Essentially, the hypotheses constitute an abstract process by which engineering designers can perform behavioral model validation. This is depicted in Figure 1.3. Designers who create a behavioral model perform validity characterization as proposed in H1 and publish the model and its validity description to a behavioral model repository or other location at which it is available to potential users. Designers who wish to conduct a simulation study search for candidate models in an appropriate repository or other suitable locations. They perform compatibility assessment to eliminate candidate models that are inappropriate in the problem context and adequacy assessment to eliminate compatible models that are not sufficiently accurate for user needs. This is the two-step assessment process proposed in H2.

The process depicted in Figure 1.3 is abstract in the sense that the hypotheses define *what* one must do but not *how* one must do it. The appropriateness of a particular method to perform one of the process steps depends on several factors, including the implementation of the behavioral model, the form of the validation-relevant knowledge and the time available for validation activities. The value of defining validation in abstract terms is that it serves as a roadmap to defining and using specific processes for specific problems. This allows one to distinguish better between the fundamental framework of concepts necessary for understanding the problem and the specific details associated with one particular solution method.

1.4.2 Hypothesis Evaluation Strategy

The strategy for evaluating the proposed hypotheses is an adaptation of the validation square (Pedersen, et al. 2000). The validation square is a four-step process for building confidence in the usefulness of a design method. Since the hypotheses proposed in this thesis do not constitute a design method, some modification of the validation square is in order. However, much of the spirit underlying the validation square is relevant for the hypotheses proposed in this thesis.

The four steps of the validation square are: theoretical structural validation, empirical structural validation, empirical performance validation and theoretical performance validation. During theoretical structural validation, one establishes the logical soundness of the method concepts individually and jointly. During empirical

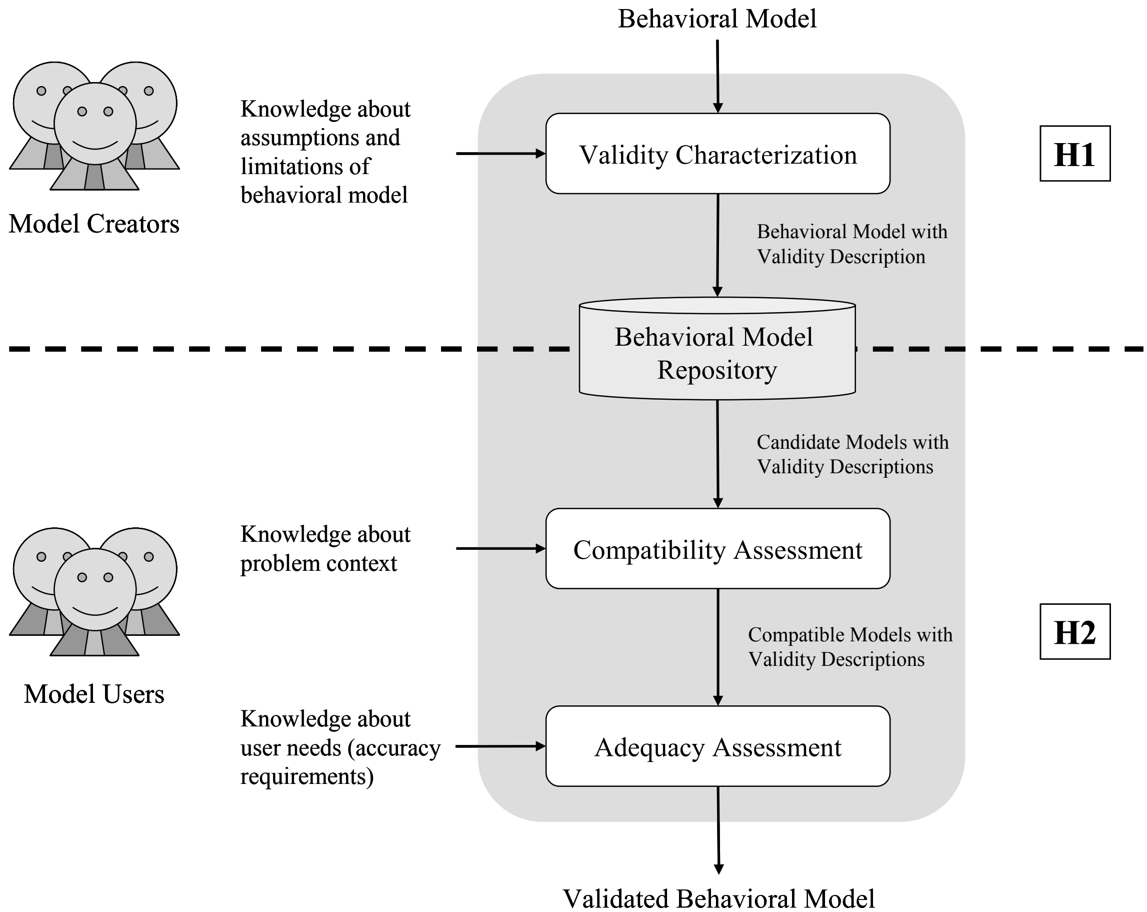

Figure 1.3: Abstract validation process and related concepts as indicated in the hypotheses.

structural validation, one establishes the appropriateness of the example problems used to evaluate the proposed method. During empirical performance validation, one gauges the performance of the method on the example problems. Finally, during theoretical performance validation, one judges the performance of the design method on problems other than the example problems.

The hypotheses proposed in this thesis do not define a specific method that can be performed on a set of example problems. Although the validation square is particularized for design methods, the underlying spirit of building confidence in the usefulness of a hypothesis is more general. This spirit is adopted in this thesis. Instead of demonstrating the usefulness of a method, the aim of evaluation efforts in this thesis is to demonstrate usefulness of a conceptual framework. A useful conceptual framework is sufficient to describe the salient features of a problem domain without being specific to any one problem or method. One should be able to specialize it to particular problems and methods easily. It should be internally consistent and consistent with the literature except for where there is a demonstrated limitation in the literature. In addition to these general criteria, the framework proposed in this thesis must accommodate the requirements stated in Section 1.3.2 that are specific to behavioral model reuse.

The main steps of the validation square are used in this thesis. The primary modification is in how the results of each step are interpreted. A description of the strategy follows. Table 1.4 is a summary of each step and the location of the corresponding material in this thesis. Figure 1.4 is a depiction of the flow of evaluation steps within the validation square construct. To improve the readability of this thesis, the

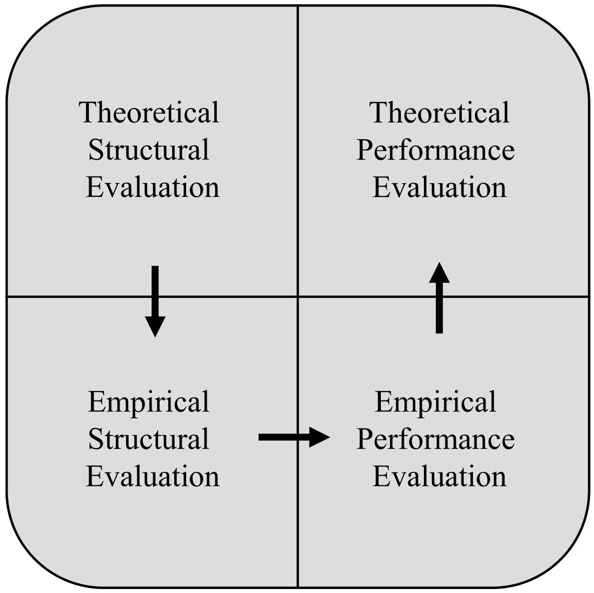

Figure 1.4: Flow of the validation square steps (Pedersen, et al. 2000).

term "validation" henceforth is reserved for the thesis subject matter. The term "evaluation" is used with respect to the process of critiquing the hypotheses

Theoretical Structural Evaluation

Theoretical structural evaluation involves establishing that the individual concepts and the overall conceptual framework are acceptable. In this thesis, acceptability of individual concepts is judged relative to whether they are appropriate for the problem domain and consistent with the relevant literature. The scope of the problem domain is described in Section 1.3. This sets the stage for discussing the suitability of individual concepts in Chapters 2 and 3. Chapter 2 contains a review of the model validation literature and establishes the suitability of the individual concepts relative to the problem

Table 1.4: Summary of hypothesis evaluation strategy for this thesis.

Evaluation Step	Description	Location
Theoretical Structural Evaluation	Review of literature to establish consistency with relevant works.	Chapter 2, Chapter 3
	Explanation of internal consistency.	Chapter 3
Empirical Structural Evaluation	Description of example problems, how the framework is particularized to solve them and what evidence about the hypotheses is provided by solving the examples this way.	Chapter 4
Empirical Performance Evaluation	Discussion of how the particularization of the framework is easily extensible for these problems.	Chapter 4
	Discussion of how other approaches to model validation are not appropriate.	Chapter 2
Theoretical Performance Evaluation	Discussion of how the conceptual framework is sufficiently rich to span the problem domain.	Chapter 5
	Discussion of how the framework applies to problems other than the examples.	
	Discussion of how the framework supports behavioral model validation methodologies that are consistent with the requirements stated in Section 1.3.2.	

domain and the relevant literature. Chapter 3 is an elaboration of the conceptual framework and further substantiates the appropriateness of the concepts.

Acceptability of the overall conceptual framework is judged relative to its internal consistency in addition to whether it is appropriate for the problem domain and compatible with relevant literature. Chapter 2 serves as support for the appropriateness of the framework relative to the problem domain and relevant literature. Chapter 3 is a substantiation that the framework is internally consistent. It also contains evidence from the literature of the appropriateness of the framework relative to the problem domain. The example problems of Chapter 4 also serve as support for the internal consistency of the framework. The basis for this is that inconsistencies likely would be exposed when solving a problem within the confines of the framework. However, this is not the primary objective behind the examples.

Empirical Structural Evaluation

Empirical structural evaluation involves establishing that the example problems are appropriate for demonstrating the performance of the conceptual framework. In this thesis, performance is judged relative to whether the framework contains the concepts necessary to describe the salient features of the validation problem and whether it is particularized to a specific problem easily. Appropriate example problems reflect problems that designers might encounter in practice. In this thesis, appropriate example problems are ones that can be solved using existing methods. This is because the hypotheses do not define any particular solution methods. To use a novel method requires evaluation of the method independent of the hypotheses. Empirical structural evaluation is performed in Chapter 4.

Empirical Performance Evaluation

Empirical performance evaluation involves establishing that the conceptual framework is useful for the example problems presented in this thesis. Often, one determines usefulness as a degree along some scale (such as the amount of cost or time reduced or improved quality of a solution). For this thesis, usefulness is a binary success or failure criteria. The basis for this is that the examples serve as proofs of concept for a proposed conceptual framework (as opposed to demonstrations of a refined solution method). Essentially, the conceptual framework is useful if the existing conceptual framework leads to methods that are inappropriate for model reuse and the proposed conceptual framework leads to methods that are viable for these scenarios.

Theoretical Performance Evaluation

Theoretical performance evaluation involves establishing that the conceptual framework is useful for problems other than the examples and that it supports methodologies that are consistent with the requirements of Section 1.3.2. This requires an examination of the degree to which the conceptual framework spans the problem domain, a discussion of how it is easily extended to other problems and a description of how it supports appropriate methodologies. Theoretical performance evaluation is performed in Chapter 5.

1.5 Thesis Organization

Figure 1.5 is a roadmap to the content of this thesis. Indicated in the figure are the objectives of each chapter. They are summarized as follows.

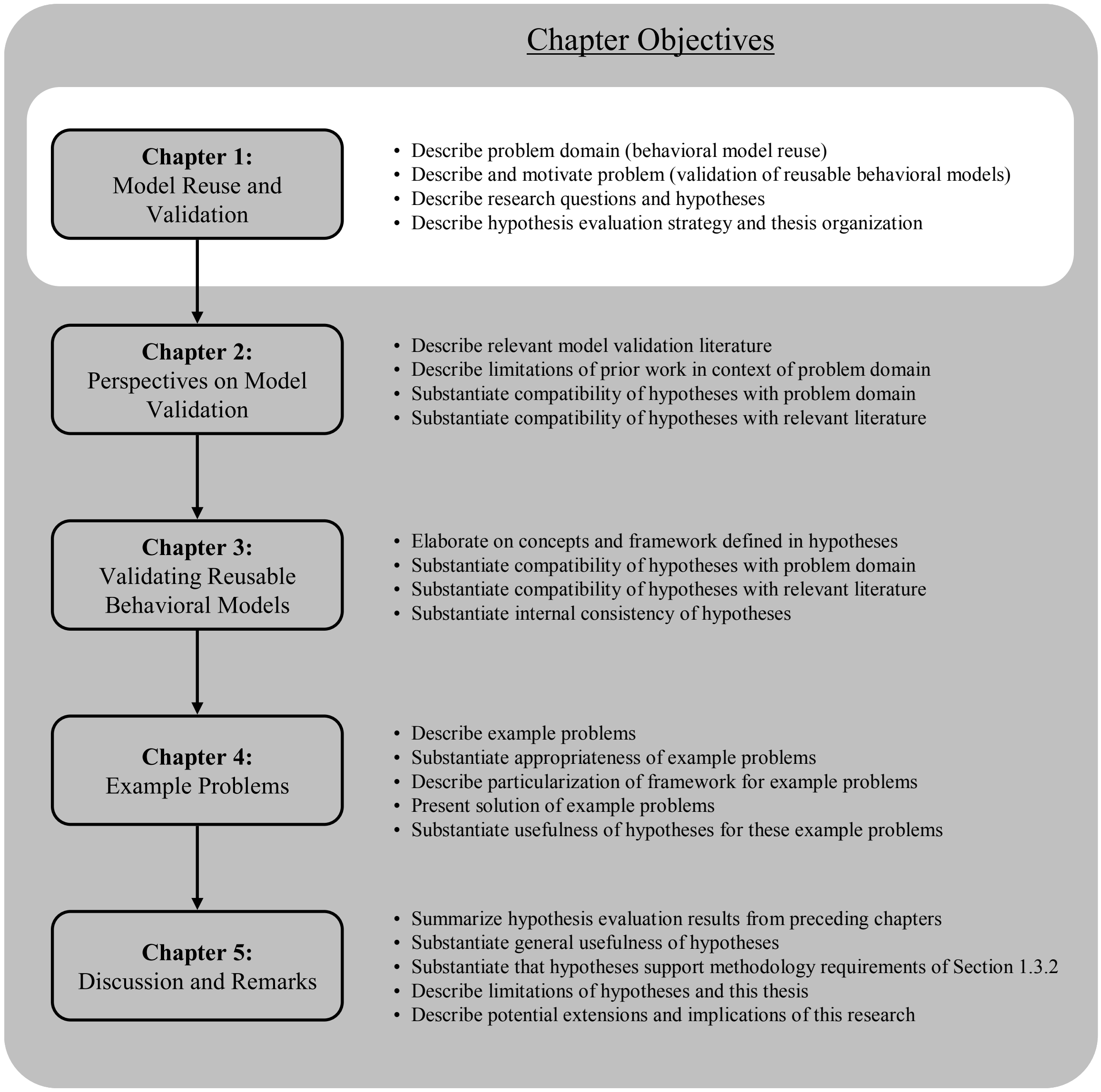

Figure 1.5: Thesis roadmap.

Chapter 1: Model Reuse and Validation

The first chapter is a description of the behavioral model reuse and the challenges of validating reusable behavioral models. It includes descriptions of specific research questions investigated and the corresponding hypotheses. It also contains a description of the strategy used to evaluate the hypotheses.

Chapter 2: Perspectives on Model Validation

The second chapter is a description of the relevant model validation literature and a substantiation of the compatibility of the individual concepts identified in the hypotheses with the problem domain and relevant literature. Thus, this chapter plays a significant role in the theoretical structural evaluation of the hypotheses.

Chapter 3: Validating Reusable Behavioral Models

The third chapter is an elaboration of the hypotheses and contains the completion of theoretical structural evaluation. Further substantiation of the compatibility of the individual concepts with the problem domain and literature is provided. The compatibility of the overall framework described in the hypotheses is substantiated. The internal consistency of the framework is substantiated as well.

Chapter 4: Example Problems

The fourth chapter contains the definition and solution of two example problems. Empirical structural evaluation is performed to build confidence in the example problems. The specific methods used to implement the proposed framework are described and the problem solutions are presented. Empirical performance validation is performed by discussing the usefulness of the framework for the example problems.

Chapter 5: Discussion and Remarks

The final chapter contains a discussion of the overall evaluation of the hypotheses and remarks about the limitations and potential extensions and implications of this research. It includes a summary of the first three hypothesis evaluation steps. This is followed by a theoretical performance evaluation of the hypotheses. Substantiation that the hypotheses are compatible with the requirements of Section 1.3.2 is presented.

CHAPTER 2:

PERSPECTIVES ON MODEL VALIDATION

This chapter is a critical review of the model validation literature and an analysis of the hypotheses in the context of this literature and behavioral model reuse. This chapter plays a significant role in the theoretical structural evaluation of the proposed hypotheses. A survey of the relevant literature provides a basis for evaluating the individual concepts of the proposed framework. Furthermore, the prevailing conceptual framework from the literature and the conceptual framework proposed in this thesis are evaluated in the context of behavioral model reuse. The prevailing framework is found to be insufficient in this problem domain, while the proposed framework is deemed appropriate. Figure 2.1 contains a summary of objectives of this chapter.

This chapter begins with an examination of model validation from a fundamental perspective. Section 2.1 is a review of accepted results from the philosophy community. This primary objective of this section is to identify the fundamental limitations and capabilities of model validation. These are general results that apply to any model validation approach. Given these fundamentals, it is appropriate to examine current model validation practice. Section 2.2 is a review of the prevailing conceptual framework and methodology for model validation. After having identified the prevailing conceptual framework, it and the framework proposed in this thesis are analyzed with respect to behavioral model reuse scenarios. Section 2.3 contains this analysis. The result is that the prevailing framework has limitations with respect to reuse scenarios, but the framework proposed in this thesis is appropriate.

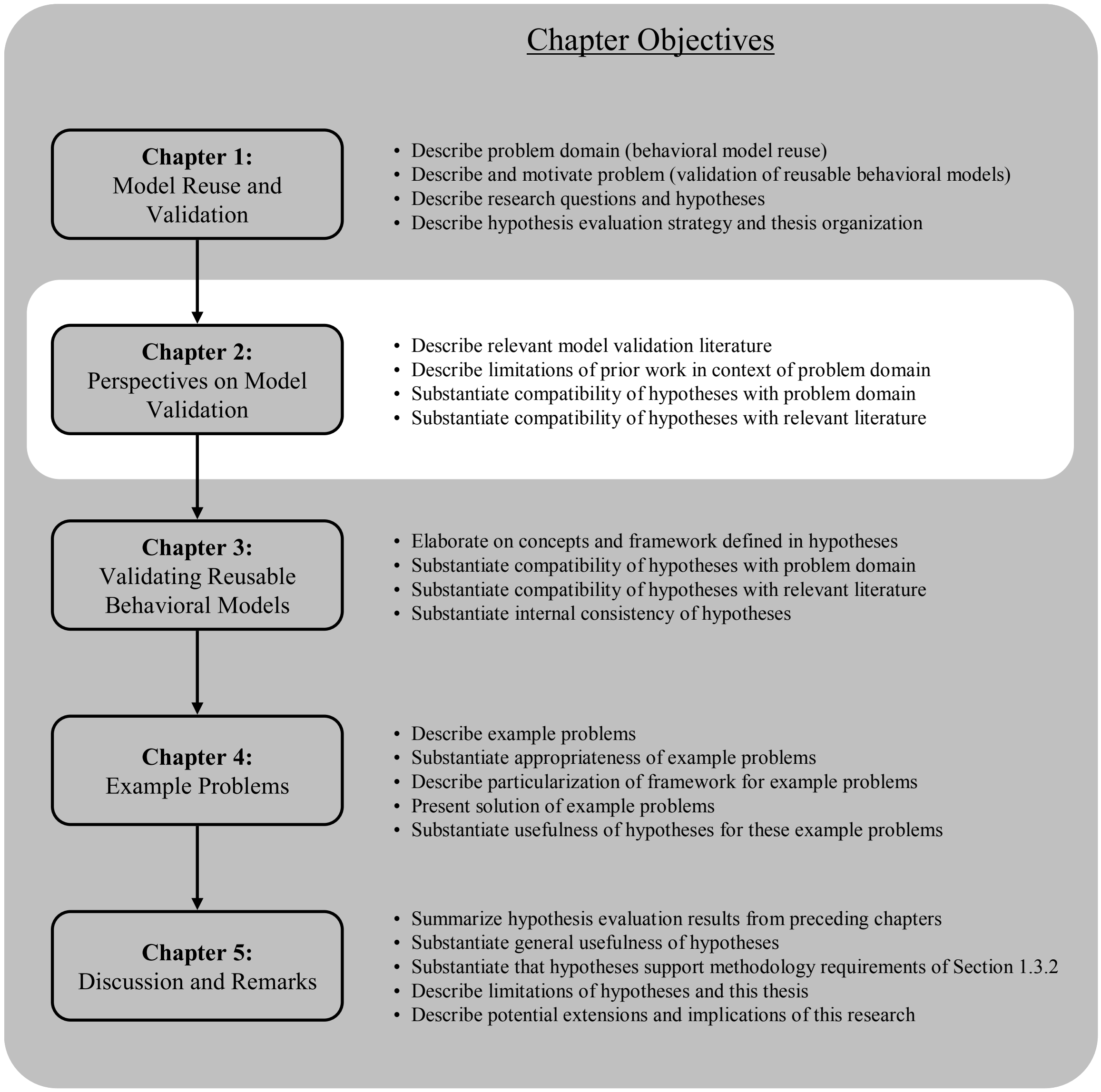

Figure 2.1: Chapter 2 objectives.

2.1 A Fundamental Perspective

In this section, the objective is to ground the discussion of behavioral model validation in a fundamental basis. This section contains a discussion of the philosophical principles that relate to model validation and the implications they have on model validation. Section 2.1.1 is an explanation of why the fundamental basis for model validation lies in philosophy and of which branches of philosophy are relevant. Section 2.1.2 is a summary of the relevant philosophical positions. Section 2.1.3 contains descriptions of several important consequences of the accepted philosophical thinking.

2.1.1 Philosophical Roots of Model Validation

Most engineering design activities have their basis in science or mathematics. For example, the fundamentals of dynamics are rooted in the laws of physics and the fundamentals of decision making are rooted in the mathematics of decision theory. These fundamentals establish the basic capabilities and limitations of activities derived from them. This knowledge is invaluable to engineers because it allows them to reason and reach sound conclusions.

Unlike many engineering design activities, model validation is not a problem of physics or pure mathematics. It is fundamentally a problem of knowledge—specifically, of determining how well something is known. Thus, the basis for model validation lies in the philosophy literature. This section is a brief survey of the relevant philosophy literature and a summary of the principal consequences for model validation. Naturally, model validation practice relies on specific methods based on mathematics, statistics and other bodies of knowledge. Any limitations due to the fundamentals of those fields are in addition to those originating from the philosophy.

Engineering designers use behavioral models to make predictions about the world. In this sense, behavioral models are similar to scientific theories. Unlike scientific theories, behavioral models do not necessarily have explanatory power. For example, a regression model may map inputs to outputs accurately relative to the real system but its structure provides no insight into why the mapping is so. None the less, it is reasonable to ask whether the problem of validating scientific theories in general is related to the problem of validating behavioral models in engineering design. In both cases, one is concerned with the relationship to the real world system. Thus, the role of observation and the strength of conclusions one can draw from observational data is paramount. The two problems differ in the way one judges the relationship to the real system. One generally considers a scientific theory to be refuted given observations that contradict it, but behavioral models may be adequate if the contradiction is within an acceptable accuracy margin. Moreover, the hypothesis associated with a behavioral model is that the model holds with a particular level of accuracy and refutation attempts are with respect to this hypothesis rather than the model itself.

Prompted by the parallel between behavioral models and scientific theories as well as questions about whether model validation can be a scientific process, several authors have explored the philosophical roots of model validation. Several authors credit Naylor and Finger (Naylor, et al. 1967) with being the first to interpret model validation in a philosophical context. More recent works have extended the dialogue by including philosophical results that were not available at the time of Naylor and Finger (e.g., (Barlas, et al. 1990, Herskovitz 1991, Kleindorfer, et al. 1993)). The article by Barlas and Carpenter is a particularly lucid survey of the relevant ideas.

The philosophical roots of model validation exist in two related areas of study, known as *epistemology* and the *philosophy of science*. Epistemology is the "study of the nature, extent and justification of knowledge" (Rosenberg 2000). The philosophy of science is the study of "questions which science cannot answer" and "questions about why the sciences cannot answer the first set of questions" (Rosenberg 2000). The boundaries between these two endeavors are not clear-cut. It is reasonable to think of the philosophy of science as dealing more closely with the nature of scientific questions and of epistemology as dealing more closely with the nature of answers to questions.

2.1.2 Epistemological Considerations

This section is a summary of the progression of different views of scientific knowledge as relevant to model validation. This is not intended as a thorough survey of the past several centuries of epistemology and the philosophy of science. The objective is to highlight the main points and ideas that have led to modern understandings of scientific knowledge and, hence, behavioral model validation.

It is convenient to organize epistemological views along a spectrum ranging from *reductionist* views to *relativist* views (Barlas, et al. 1990). Readers should be aware that others describe this spectrum as justificationist-to-antijusificationist (e.g., (Kleindorfer, et al. 1993)) and that many other near-synonymous characterizations are possible. Reductionists tend to envision knowledge as objective "truths" that await discovery, while relativists tend to see knowledge as more subjective and open to interpretation. The degree to which a particular belief system is reductionist or relativist is debatable.

However, it is usually possible to agree that a particular viewpoint is closer to one end of the spectrum than the other.

<u>Reductionism</u>

Reductionist views, including those described as *foundationalist*, *formalist*, *justificationist* or *logical positivist*, maintain that knowledge is objective and it has a basis in observation (the *empiricists*) or self-evident truth (the *rationalists*) (Kleindorfer, et al. 1993). Such views hold that there must be some justification in favor of considering something to be scientific knowledge. Moreover, hypotheses must somehow "pass tests" in order to be confirmed as scientific knowledge (Miller 1994). Descartes (a rationalist) and Locke (an empiricist) were among the classical philosophers subscribing to such beliefs. While particular incarnations of reductionism vary, they all assume a pure objective truth exists and can be discovered. These views are aptly described as utopian—a quest for the elegant, objective and logical universe.

Logical positivism (also called *logical empiricism*) was an early movement in the philosophy of science that is strongly reductionist. They sought a rigorous framework by which scientists would deductively prove unambiguous statements (Rosenberg 2000). Their goal was to establish a sound basis for discovering scientific truth, based upon a foundation of objective observational facts (Barlas, et al. 1990). Believers in logical positivism hold that concepts are scientifically meaningless absent direct observation of the item or property for which it is named. This led some logical positivists of the late-19th and early-20th centuries to deny the meaningfulness of concepts such as "atom" or "molecule" on the basis that there was no direct empirical evidence of them (Rosenberg 2000).

Challenges to Reductionism

One early challenge to reductionism in science was the problem of *induction*. David Hume originally described the problem in *A Treatise of Human Nature* (Hume 1965). The problem is that of generalizing beyond observational data. Strict empiricists hold that all knowledge results from logical reasoning about observational data. Observations of the world are necessarily finite, yet most scientific theories make predictions over continuous space and time. Hume pointed out that one requires inductive inference to verify such theories from empirical data, but that there can be no empirical basis for induction. Thus, he showed a contradiction in the pure empiricist epistemology.

Karl Popper introduced the notion of *falsification* as a response to the problem of induction (Popper 1972). His position was that theories may be proven false, but may never be proven true. Instead, all theories are provisional and exist only until a counterexample discredits it. Previously, philosophers of science sought to find *justifications* for a statement to be considered knowledge (typically by logical inference). As Miller puts it, falsification "relies on expulsion procedures, rather than entrance requirements" (Miller 1994). This contrasts with the justificationist views that tend to coexist with reductionism. Falsification also solves the problem of identifying scientific theories, called the problem of *demarcation*. According to Popper, a scientific theory must be falsifiable based upon empirical evidence (Popper 1983).

Another problem with reductionist epistemologies is their foundation on the presumed existence of observational objectivity. Many reductionists, particularly empiricists and logical positivists, believe that one can observe the world independently of any bias or prior belief. This view is largely discredited by a number of works dating from the middle-twentieth century. Notably, Kuhn argues that theory-free observation is not possible and proposes a view of science in which observations are interpreted according to the reigning "paradigm" of the day (Kuhn 1996). He presents several historical examples in which immature sciences (i.e., those that do not have a reigning paradigm) are comprised of several "schools" with differing theories about a field. These "schools" seldom communicate results because the interpretation of the results is so closely tied to the theory within which they are interpreted. The discrediting of empirical objectivity effectively leaves reductionist views defeated. Although in its pure form falsification is reductionist in terms of its reliance on objective observation, modern interpretations preserve the idea that theories may only be refuted while substituting modern understandings of what constitutes evidence.

Relativism

Relativist views, including those called *holistic*, *social* or *antijustificationist*, acknowledge the failure of a purely objective and logical pursuit of science (and knowledge in general) (Barlas, et al. 1990, Kleindorfer, et al. 1993). There exist many epistemologies that may be classified as relativist and some may have conflicting premises. The common thread among them is their acceptance of the social, historical and cultural biases in observation and interpretation. They agree that pursuit of "the truth" is fruitless and favor more pragmatic approaches. Although some relativist views have strayed far from the central issues of the philosophy of science (as noted by (Rosenberg 2000) and (Miller 1994)), most such perspectives do not necessarily deny the roles of logic and sound reasoning and the predictive value of scientific knowledge. They merely admit the intrinsic limitations of such methods.

Broadly, relativism represents the most recent thinking in epistemology and the philosophy of science. These views came to the forefront of philosophical thinking only in the latter part of the twentieth century. There is no reason to presume that these views are the final answer in the debate, though. Interestingly, the very principle of relativist belief—that knowledge is relative to culture and history—implies that our understanding of knowledge itself will change over time.

2.1.3 Consequences for Model Validation

The results of philosophy have several consequences for model validation. The foremost of these are described below. It is worth noting that some might consider these conclusions "self-evident" or "common sense" to an average scientist or engineer and therefore not requiring a basis in the philosophy. However, this perspective ignores the fact that the fundamental philosophical understandings that lead to these "obvious" statements are distilled throughout the modern engineering education. Although they may not always be stated explicitly, these philosophical understandings of knowledge are the thread from which the fabric of scientific and engineering understanding are woven. An explicit articulation of the ties between the philosophy and its consequences for model validation has two benefits: First, it introduces an element of traceability into the community's understanding of model validation. Should understandings of knowledge shift, this linkage between model validation and philosophy can serve as a roadmap for reexamining model validation according to the new epistemology. Second, it establishes the capabilities and limitations of model validation on firm ground. It is important to remember that statements that are "self-evident" to one person may not be so to another. An articulation of fundamentals provides a basis for unambiguous dialog.

No amount of data "proves" a model to be valid.

This is a direct consequence of the problem of induction. To prove deductively that a continuous model is valid, one would require an infinite amount of empirical data. To consider a model valid without an infinite amount of data requires a "leap of faith" on the part of its user. Balci reaches a related conclusion but from a practical perspective (Balci 1997). He states that "complete simulation model testing is not possible." He justifies this by arguing that the number of possible input combinations that one must test for a model of reasonable complexity is so large as to be impossible to do within realistic budgetary and time constraints. This is essentially a pragmatic slant on the problem of induction.

Validation is necessarily subjective.

This is a consequence of induction and the relativist perspective on observation. The problem of induction implies that models are *assumed* true rather than *proven* true. The basis for such assumptions ultimately is subjective; for example, one person may be satisfied after seeing 10 data points, while another wants to see 100. The relativist perspective on observation holds that all observations are biased by preexisting theory or belief. Thus, even if induction were not an issue, there would be a question of how one interprets the measurements. This raises questions about how validation can proceed in a collaborative environment. Can a designer draw a reasonable (subjective) conclusion based upon the subjective conclusions of other designers? How can collaborating designers ensure they employ compatible interpretations of results? This sentiment is echoed in the experiential work of Balci (Balci 1997). He states that validation "requires independence to prevent developer's bias." His statement is a consequence of inherent subjectivity and a presumed (typically inadvertent) bias on the part of model developers to be “convinced” of the validity of their model more easily than would be an independent arbiter.

<u>Validation can be scientific.</u>

Falsification holds that a statement is scientific if it can be refuted. Thus, validation is a scientific endeavor if statements of validity are refutable. While the basis for refutation may be subjective, the scientific process can proceed as long as the participants can agree on what constitutes a conclusive refutation of a statement. The participants should also agree on what constitutes a sufficient attempt to refute a statement (i.e., when to assume it to be valid due to lack of refutation). This implies that behavioral models, as scientific theories, have predictive value if they survive reasonable falsification attempts.

<u>Valid statements can be deduced from other valid statements.</u>

Once an item is accepted as valid by a person, that person can use it in deductive reasoning. However, deduction is a “garbage-in, garbage-out” process and use of one invalid premise or rule invalidates all subsequent conclusions. In this sense, validation in engineering design is a vetting process that weeds out invalid premises and rules and that allows design to proceed in a deductive and accountable fashion.

<u>Validation requires a basis for trust when multiple people are involved.</u>

Validation is the pursuit of trust and to achieve it requires some base level of trust among the participants. This conclusion is a corollary of those above. Since validation cannot be “proved” and is not objective, the various parties involved with validation activities must have some basis for trusting one another. This basis manifests itself in terms of accepted procedures for gathering, interpreting and reporting data and in terms of standards for scientific inquiry. Typically, this basis is implicit and arises through the commonalities in our education as scientists and engineers. However, the stronger and more explicit is the fundamental basis for trust, the stronger are the conclusions about validity that one can draw. At a minimum, an explicit basis for trust improves process traceability. This is particularly important when model reuse is an issue because model reuse typically involves multiple participants. This is evidenced in the model reuse scenarios of Section 1.3.1.

2.2 A Practical Perspective

The previous section is an account of the fundamental limitations and capabilities of model validation. The objective of this section is to describe the practical response to these limitations that is proposed in the literature. Section 2.2.1 is a description of the prevailing view of validation as a confidence-building activity. It includes a description of the related issue of verification. Section 2.2.2 is an account of a conceptual framework for model validation that is based on a simplified modeling and simulation process.

2.2.1 Confidence-Building through Verification and Validation

A "valid model" is sufficiently accurate for a user's needs over the set of intended model scenarios. However, a main consequence of the prevailing philosophical thought is that, due to the problem of induction, it is impossible to prove the accuracy properties of a model. This raises an important question: *if it is impossible to prove validity, how should model developers evaluate their models?* The prevailing view in the literature is that model evaluation is a confidence-building activity (Knepell, et al. 1993, Balci 1997, Law, et al. 2000). The essential position is that one cannot prove validity, but *one can perform activities that increase confidence in the validity of a model*. This perspective is consistent with the fundamentals. To conclude that a model is valid requires a leap of faith, but one need not make this leap without any evidence.

Accuracy assessment is a key factor in confidence building. Two basic processes that yield evidence about model accuracy are *verification* and *validation*. Validation processes include activities that gauge the accuracy of a model *relative to the system it represents*. For example, model validation might involve a comparison of model outputs to system outputs for corresponding inputs. One performs accuracy assessments in the context of the intended model uses and judges accuracy with respect to user needs. However, appropriate definition of intended model uses and user needs is generally considered beyond the scope of model validation. Together with model validation and other accuracy-assessment processes, these steps address the problem of *model credibility assessment* (Balci 1997). Although important in the pursuit of confidence in simulation results, these steps are not a focus of this thesis.

Verification processes include activities that gauge the accuracy of a model *relative to another model*. This type of accuracy assessment is important in the modeling process, which generally is iterative and involves transformations from one representation to another. For example, developers often must transform a set of mathematical equations into a format amenable to computer-numeric analysis. Verification is a process by which developers confirm they have performed such transformations properly. This allows developers to gauge their progress throughout the modeling and simulation process.

Verification also serves as a mechanism for building confidence in model validity. This is possible when a developer transforms a previously validated model into another representation. The developer has already established evidence about the accuracy of the first model relative to the system it represents. If both models represent the same system—say the first model is a set of equations and the second is a computer-interpretable implementation of the equations—then verification is a link between the transformed model and the system. The developer can gauge the accuracy of the second model relative to the first by verification and the accuracy relative to the system by association with the first model. This is an associative relationship. Informally, the reasoning is: "if my first model was right relative to reality and my second model was right relative to my first, then my second is right relative to reality."

The terms verification and validation are synonyms in everyday usage that academics appropriated for specialized use. One way to remember the distinction between the words as used in academics is to consider the questions that they address (Balci 1995):

<u>Verification</u>: Did we build the model right?

<u>Validation</u>: Did we build the right model?

Verification relates to whether modelers correctly implement their intent, while validation relates to whether their intent is appropriate in the first place. Another way of looking at this is to consider which entities are related by each process. From this perspective:

<u>Verification</u>: Relates one model to another.

Validation: Relates a model to a system.

Verification can serve to propagate knowledge about model validity, but it cannot establish this knowledge because it does not involve a relationship to the system being modeled.

2.2.2 A Framework for Model Validation

Figure 2.2 is an illustration of a conceptual framework for performing model validation in a modeling and simulation process. It is useful for understanding how one can use verification and validation to build confidence in a model. The figure is an abstract depiction of the modeling and simulation process and relates various verification and validation processes to the modeling and simulation steps. It is based on the work of Sargent (Sargent 1985, 1987), whose work draws from earlier work by the Society for Computer Simulation Technical Committee on Model Credibility (Schlesinger, et al. 1979).

Other frameworks with the same objectives are described in the literature. These are fundamentally similar to the framework of Figure 2.2, but are defined in terms of more elaborate modeling and simulation processes. Banks, Gerstein and Searles survey several of these, including the one described here (Banks, et al. 1987). They find that the abstract process includes all of the fundamental aspects of validation and verification and that it is particularly clear and understandable. Thus, this discussion is limited to the framework based on the abstract process.

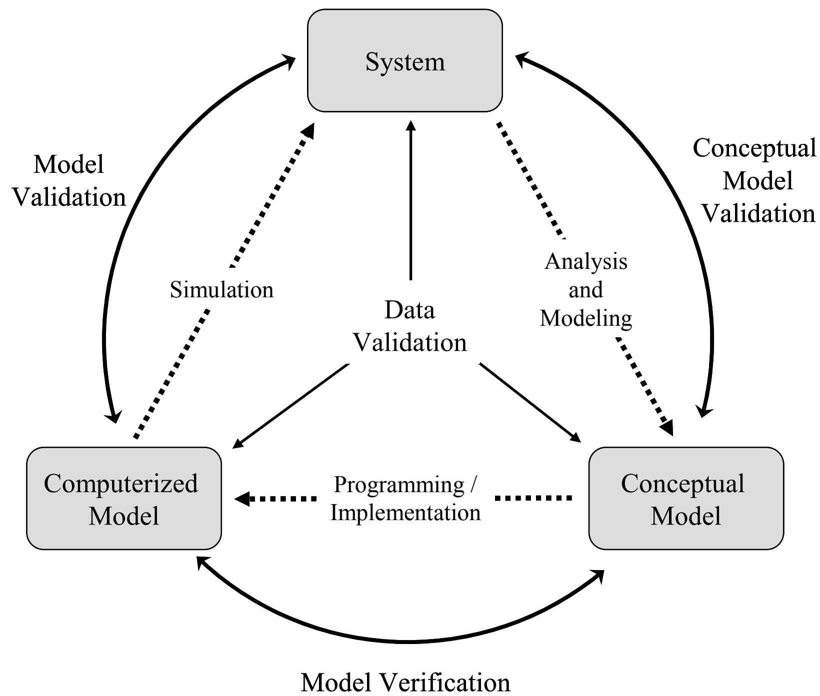

Figure 2.2: A model validation framework based on an abstract modeling and simulation process.

System, Conceptual Model and Computerized Model

A *system* is something, existing or proposed, that is to be modeled. In the context of engineering design, this typically is a design alternative. It can refer to both an entity (e.g., a beam) or an entity in a particular context (e.g., an axially loaded beam). Some authors use other terms with the same meaning. Examples include “problem entity” (Sargent 1985, 1987), “real world” (Robinson 1997), “reality” (Schlesinger, et al. 1979) and “substantive problem” (US GAO1979). System is preferred in this thesis because it has more meaning in the context of engineering design.

Conceptual models and computerized models are different states of maturity in the model development process. A *conceptual model* is a mathematical, logical, verbal or mental representation of the system. A *computerized model* is a computer-interpretable implementation of a conceptual model. It is produced through a programming and implementation process. Computerized models are what model users evaluate during a simulation study. They sometimes are referred to as “simulation models.” Conceptual models are an intermediate form that one cannot directly evaluate on a computer. Unless specified otherwise, use of the term “behavioral model” in this thesis refers to a computerized model.

Model developers arrive at a conceptual model through a modeling and analysis process. A conceptual model reflects all of the salient features of the system it represents. This includes any assumptions or phenomena assumed relevant to the simulation study. One typically develops a conceptual model in an iterative fashion. It may begin as a mental or verbal representation and become logical or mathematical as it matures. More sophisticated conceptual frameworks for modeling and simulation address this progression explicitly (e.g., see (Nance 1984, Banks, et al. 1987)).

Validation and Verification Processes

The validation and verification processes are defined below.

> <u>Data Validation</u>: The process of ensuring that the data necessary for M&S activities (model building, evaluation and testing, simulation, etc.) are adequate and correct (Sargent 1985).
>
> <u>Conceptual Model Validation</u>: The process of substantiating that the theories and assumptions underlying a conceptual model are correct and

that the representation of the system is reasonable for the intended purpose of the model (Sargent 1985).

Model Verification: The process of assuring that a computerized model is a sufficiently accurate representation of the corresponding conceptual model (Schlesinger, et al. 1979, Sargent 1985).

Model Validation: The process of substantiating that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model (Schlesinger, et al. 1979).

Data validation is central to all other validation and verification activities as well as the model development process. In this setting, data refers to numerical results of experiments. There are three purposes for data in a modeling and simulation study: for developing the conceptual model, for validating the computerized model and for performing experiments with the validated model (Sargent 1987). Incorrect or inappropriate data can undermine any of these activities. Specific data validation considerations include ensuring that data is accurate and consistent. Data accuracy depends on several factors, including how it was acquired (e.g., measurements, predictions from other models, etc.) and how it is treated after acquisition (e.g., how it is represented, transformed, etc.). Inconsistent data (e.g., sets of data that suggest mutually exclusive conclusions) must be investigated and resolved. This can involve tracing data back to their source and possibly repeating experiments from which they were acquired. In general, data validation requires knowledge of the data source and its use. It typically is addressed by enforcing specific gathering and analysis procedures during the course of a project.

Conceptual model validation sometimes is referred to as “theoretical validation” (US GAO1979) or “model qualification” (Schlesinger, et al. 1979). Because the development of conceptual models is iterative, conceptual model validation is an ongoing process. For each modification and refinement made to a conceptual model, developers must consider its impact to validity. The focus is on rationalizing the structure and assumptions embodied in a model. This involves in-depth knowledge about the conceptual model, the system being represented and the intended model scenarios. The process is driven by expert opinion and analysis and supplemented by observational data when possible. Several authors describe methods and approaches to conceptual model validation. One approach consists of five primary activities: face validity analysis; historical analysis; intended use and requirements analysis; model concepts and fidelity analysis; and logic trace analysis (Knepell, et al. 1993). Each of these activities is highly dependent on manual review, expert analysis and systematic documentation procedures. An approach involving document-based communication among model developers is recommended in a survey paper (Robinson 1997). Another survey identifies seven methods suitable for conceptual model validation, all of which are informal and expert-based (Balci 1998). Sargent notes that when possible one should use observational data to substantiate modeling assumptions (Sargent 1987).

According to this framework, *model verification* relates a conceptual model to a computerized model. One may also interpret it more broadly to apply to any model transformation. For example, one can verify the transformation of a verbal model description into a mathematical description or of one mathematical description into another. The latter case is common when simplifying a model from a form that is highly accurate but difficult to evaluate. The difficulty of model verification depends on the technologies being used. It is a significantly more difficult task when the model is implemented using a programming language such as C or Java. In these cases, developers typically must implement support code—as equation solvers, user interfaces, etc.—in addition to the model itself. Developers typically turn to software engineering methodologies for guidance on such tasks. One methodology involves various software tests, such as program logic analysis, program interface analysis and program constraint analysis (Knepell, et al. 1993). Other authors present analogous methods (e.g., (Robinson 1997, Balci 1998, Oberkampf, et al. 2002)). Recent advances in modeling and simulation tools are changing the way that developers can implement their models. Modeling languages like Modelica (Mattsson, et al. 1998, Tiller 2001) and VHDL-AMS (Christen, et al. 1999) allow implementation with a nearly one-to-one mapping between a conceptual model and the modeling language description (i.e., the computerized model). This reduces model verification to a nearly trivial task.

As described in the context of this framework, *model validation* is substantially similar to the notion of behavioral model validation defined in Chapter 1. The concept also is known as "operational validation" (Sargent 1985). At a notional level, one can consider model validation and conceptual model validation to be the same since they both involve assessing the accuracy of a model relative to reality. The main distinction between them is a practical one. Conceptual model validation is more of a "white-box" problem—one considers the appropriateness of underlying structure and assumptions. In contrast, model validation is more of a "black-box" problem in which one deals primarily with the input-output relationship implemented by a computerized model as compared to the input-output relationship of the system it represents. One typically is more concerned with whether a model makes sufficiently accurate predictions than with whether the implementation of the model reflects the principles at work in the real system. For instance, the structure of a regression model is unrelated to the principles that govern a system it represents but it can perform well from a model validation standpoint.

Methods for Assessing Model Accuracy

The means by which one decides whether a model has a "satisfactory range of accuracy" vary throughout the literature. Methods range from visual inspection of data to formal inference and statistical hypothesis testing (Sargent 1985, Balci 1994, Balci 1995). Visual inspection methods involve expert opinion about what constitutes "satisfactory." A common approach, called the face validity technique, is for an expert to observe model predictions and system responses on the same axes and to decide whether they are sufficiently similar for user needs (Sargent 1985, Knepell, et al. 1993). Another approach is to perform a Turing Test in which an expert views graphs from the model and system independently and tries to identify which is which (Schruben 1980, Sargent 1985, Knepell, et al. 1993). The premise is that if experts cannot distinguish between the graphs, then the model is adequate. Other visualization techniques also provide opportunities for expert analysis. For example, experts can examine animations of model results. This approach is appropriate for evaluating complex, time-dependent relationships and is particularly useful for identifying errors (e.g., a robot arm intersecting itself, a projectile "falling" up, etc.) (Knepell, et al. 1993).

Often, visual techniques are insufficient for determining whether model accuracy is satisfactory. Because they are relatively fast and easy to perform, it is common practice to use visual techniques as a first check and then follow them with more discerning methods (Knepell, et al. 1993). The more discerning methods, sometimes called analytical methods or analytical tests, rely on formal mathematical procedures to compare simulation results to system responses. These methods include various types of confidence interval analysis and hypothesis testing. Their primary limitation is that they require a relatively large amount of data. This can be problematic when system data is scarce. It often is prohibitively expensive or time consuming to gather data of the system in the intended modeling scenarios. In such situations, model validation must proceed with minimal data. Other times, observational data is nonexistent. This can happen when it is impossible or unacceptably dangerous to perform the appropriate experiments on the system or when the system does not exist (as is often the case for problems related to engineering design).

In lieu of comparing results from the computerized model to observations of the system, some authors argue that one can perform model validation by comparing results from the computerized model to results from another model that is accepted as "true" or "valid" (Knepell, et al. 1993, Robinson 1997). One typically uses results from the accepted model as though they are produced by the system. This approach is common when the accepted model is a higher-fidelity representation of the same system. Birta and Ozmizrak describe an approach in which one compares computerized model results to an expert system that contains rules that describe appropriate input-output relationships for the system (Birta, et al. 1996). In their approach the accepted model is the knowledge base. Such approaches build confidence in the validity of a computerized model by relating it to a model of established validity characteristics.

2.3 A Model Reuse Perspective

The previous section contains a description of the prevailing conceptual framework for model validation and the approaches based upon it. This framework reflects the fundamentals described in Section 2.1 and is rooted on an abstract conceptualization of the modeling and simulation process. The objective of this section is to examine model validation from a model reuse perspective. Section 2.3.1 is an explanation of the inappropriateness of existing model validation approaches in situations of behavioral model reuse. In Section 3.3.2 it is argued that the hypotheses proposed in Chapter 1 are an appropriate framework for the validation of reusable behavioral models.

2.3.1 Limitations of Existing Approaches to Model Validation

Although appropriate for many problems, model validation approaches based on the framework depicted in Figure 2.2 are inappropriate for many behavioral model reuse scenarios. The main limitation of the framework is that it lacks concepts for describing some situations that arise during model reuse. It includes no mechanism for describing validation-relevant knowledge. Furthermore, it includes no distinction between model creators and model users and, as a result, no way to describe their interactions. This lack of expressiveness is an artifact of the modeling and simulation process upon which the framework is based. Concepts such as validation-relevant knowledge are unnecessary in this process and therefore not a part of the model validation framework.

The process underlying the framework of Figure 2.2 is appropriate for scenarios in which the use of a model is known at the time of model development. The identification of one "System" entity in Figure 2.2 is an indication of this. Without this assumption, one could not build confidence that a model is valid with respect to its use through conceptual model validation and model verification. Another assumption underlying the process is that model creators and model users interact freely, sharing knowledge as needed. Because of this assumption, there is no need for one to distinguish between the knowledge of a model creator and that of a model user. It also means that model validation can be an ongoing process throughout model development.

In general, model reuse involves independent development and use processes. One possible arrangement is depicted in Figure 2.3 (Mocko, et al. 2004). This arrangement is based on the reuse of behavioral models within an engineering design process. During behavioral model development, model creators develop a model and publish it to a suitable repository for reuse. The model use process is an alternative evaluation task that is part of a larger design process. Model users (engineering designers) begin by formulating an evaluation problem. Based upon this problem formulation, they search a repository and select an adequate model from the available candidates. It is during this step that users perform model validation—they must only select models they are confident to be valid in the context of their evaluation problem. The final steps involve simulation and evaluation of the results in the context of the engineering problem.

Because model reuse involves independent development and use processes, it is incompatible with the assumptions that creators know the details of model use prior to development and that creators and users interact closely. The first assumption

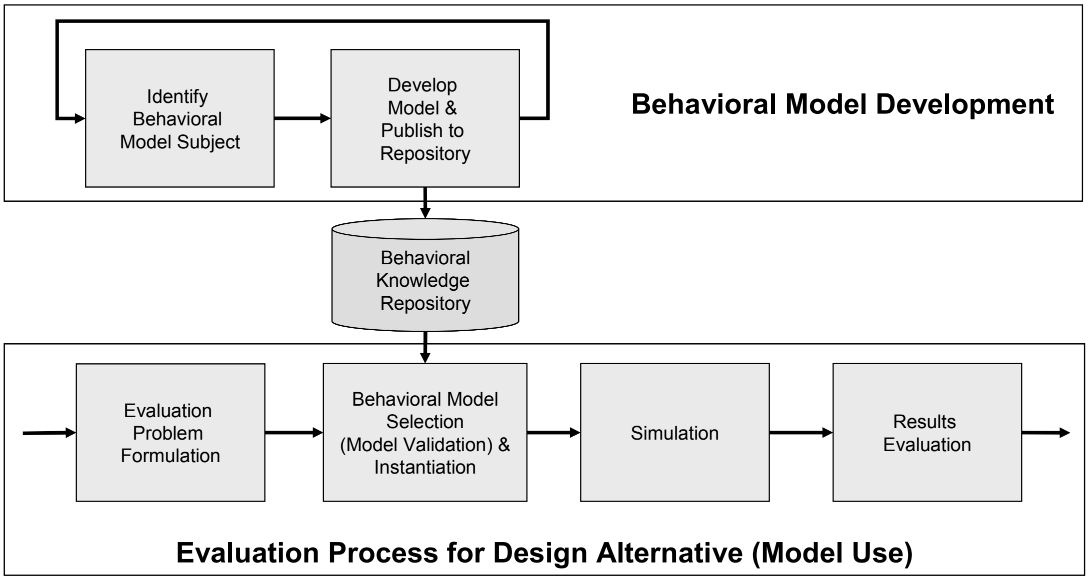

Figure 2.3: Relationship between model development and model use processes for behavioral model reuse in engineering design (Mocko, et al. 2004)

does not hold by definition. Model reuse involves the application of an existing model in lieu of developing a new model. Essentially, model creators cannot foresee all future uses of the models they develop. The second assumption holds sometimes, but not in general. Reasons for this are cited in Section 1.3.1. For example, a model creator could go to work for a competing company and therefore no longer be available for consultation. In some reuse scenarios, the only communication between model creators and model users occurs via the repository contents.

Figure 2.4 is an illustration of a model validation problem corresponding to the reuse process of Figure 2.3. Model creators develop a computerized model and publish it to a repository. They develop it with respect to a particular system, System A, and may

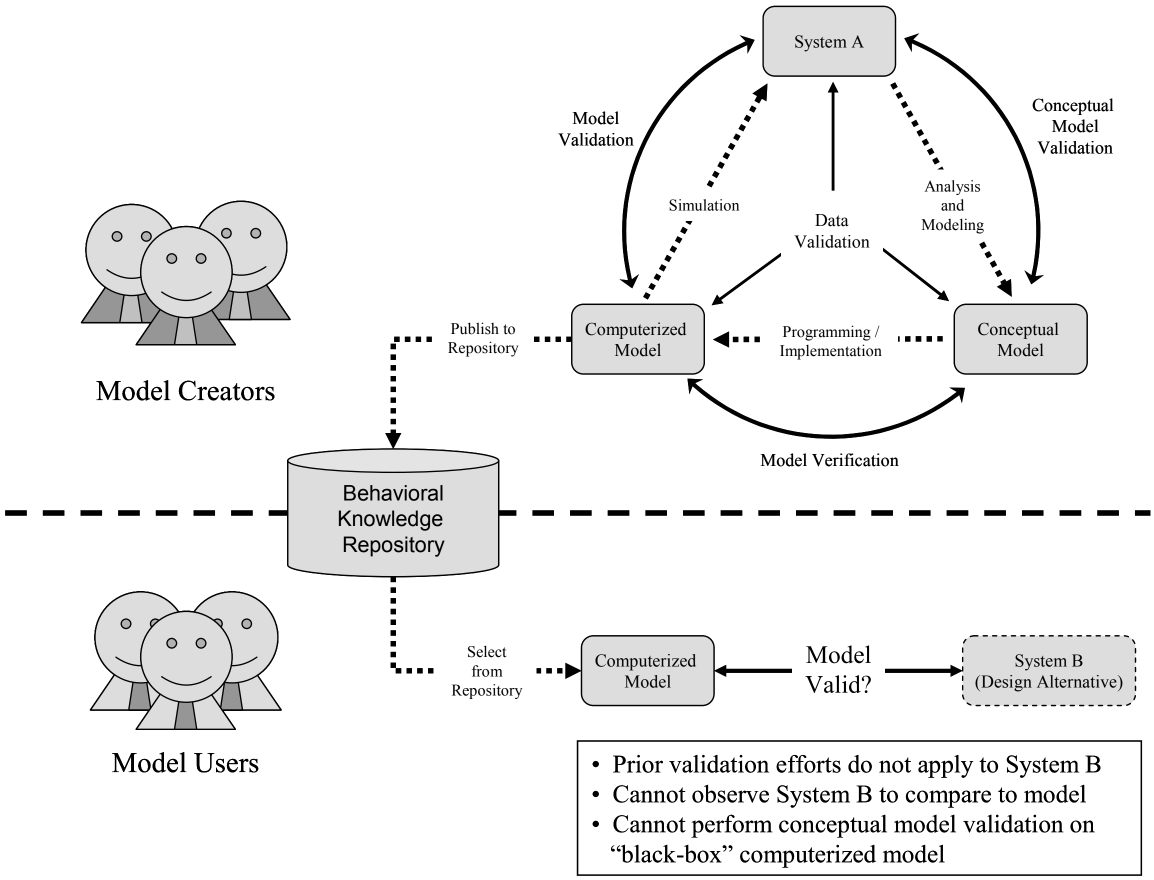

Figure 2.4: A depiction of the model validation problem for instances of behavioral model reuse in engineering design.

perform model validation relative to that system. In a model reuse situation, the objective of model users is to evaluate the behavior of a system, System B, with a preexisting model that is valid for their needs. However, model validation is more complicated in a reuse scenario in engineering design than in scenarios compatible with Figure 2.2. The availability of validation-relevant knowledge is a key consideration:

- Model creators cannot build confidence in the validity of a model relative to System B because they do not know its details during model development. They can build confidence in the validity of a model relative to a known system, System A, using approaches based on the framework of Figure 2.2. However, this confidence does not generally transfer to a different system, System B.
- Often for engineering design problems, System B corresponds to a design alternative that does not exist. A major motivation for conducting a simulation study is to evaluate the alternative without having to build it. However, this prevents model users from gauging the accuracy of a model through direct comparison to the system. Thus, users must build confidence through conceptual model validation.
- Model users know the details of the target system, System B, but lack knowledge about the underlying assumptions of the model. Without this knowledge, model users cannot build confidence in model validity relative to System B through conceptual model validation.

Thus, model creators lack knowledge about System B while model users lack knowledge about the assumptions and limitations of the model. Validation-relevant knowledge is a focal point of building confidence in reusable models. Any model validation approach that is appropriate in reuse scenarios must include means to describe and solve this problem. Any appropriate conceptual framework must include concepts and relationships sufficient to describe this problem.

2.3.2 Appropriateness of the Hypotheses

Given the deficiencies of the prior work on model validation with respect to model reuse, it is evident that a new approach is needed. What is more, a new *conceptual framework* is needed that is suitable for the problem of validating reusable behavioral models. The hypotheses proposed in the first chapter comprise such a conceptual framework. The following is an explanation of why the hypotheses are appropriate for establishing the validity of models in reuse scenarios.

The principal problem described in Section 2.3.1 is the separation of validation relevant knowledge from where it is useful. In terms of Figure 2.4, model developers lack knowledge about the system and its use while model users lack knowledge about the properties of the model and had no avenue for gathering observational data about the system. The first hypothesis addresses how creators and users should communicate validation-relevant knowledge. It is repeated below.

> ***H1:*** *Model creators can develop mathematical descriptions of their creations—called validity descriptions—that provide assertions about the accuracy a user can expect and the context over which the assertions hold true.*

The hypothesis calls for model creators to provide users with descriptions of behavioral models that are relevant during validation. This step can bridge the divide in knowledge between creators and users. The following definition for behavioral model validation is provided in Chapter 1:

> Behavioral Model Validation: The process of determining whether a behavioral model is sufficiently accurate for user needs in the context of intended model scenarios.

Model creators cannot foresee particular user needs and model use contexts and therefore cannot perform build confidence in the validity of a model for general reuse scenarios. For users to provide this knowledge to creators prior to development would not be a reuse scenario. Furthermore, it is difficult for model users to acquire the validation-relevant knowledge they lack on their own. Because they typically lack knowledge of the assumptions and limitations of a model, they cannot quantify model accuracy on a theoretical basis. Because they typically lack a physical system against which to compare a model, they cannot quantify accuracy through empirical means. The only acceptable flow of validation-relevant knowledge in a reuse scenario is *from model creators to model users*.

Validity descriptions contain the validation-relevant knowledge that model creators can acquire and that model users require. Model creators can quantify the accuracy of a model in a particular context within the confines of the framework of Figure 2.2. Rather than concluding whether a model is a sufficiently accurate representation of a particular system over a particular context, they quantify the accuracy of the model in some context. Because accuracy can depend on context, model creators must report both the quantified accuracy and the context in which they made the quantification. By formulating validity descriptions in mathematical form, creators provide model users with an unambiguous account of the validation-relevant properties of a model. This is particularly important when model creators may not be available to provide clarifications at the time of model use.

The second hypothesis addresses how model users apply validation-relevant knowledge communicated to them by model creators. The hypothesis is repeated below:

> ***H2:*** *Model users can perform a two-step assessment process in which they:*
>
> *(1) determine whether the context stated in the validity description is compatible with the problem and, if it does,*
>
> *(2) determine whether the accuracy stated in the validity description is sufficient for the needs of the problem.*

Given the knowledge codified in a validity description, model users can apply their knowledge about their accuracy requirements and intended model user context to reach a conclusion about the validity of a model with respect to their simulation problem. The first step of H2, compatibility assessment, involves a comparison of the intended use context of the simulation study and the context from the model's validity description. The second step of H2, adequacy assessment, involves a comparison of the user's accuracy requirements and the model accuracy in the stated context. Together, these establish whether a model is sufficiently accurate for user needs within the context of intended model use.

Figure 2.5 is an illustration of a model validation problem corresponding to the reuse process of Figure 2.3 that includes the concepts from the hypotheses. The model development process includes validity characterization performed within the framework of Figure 2.2. Model creators publish the resulting computerized model and associated

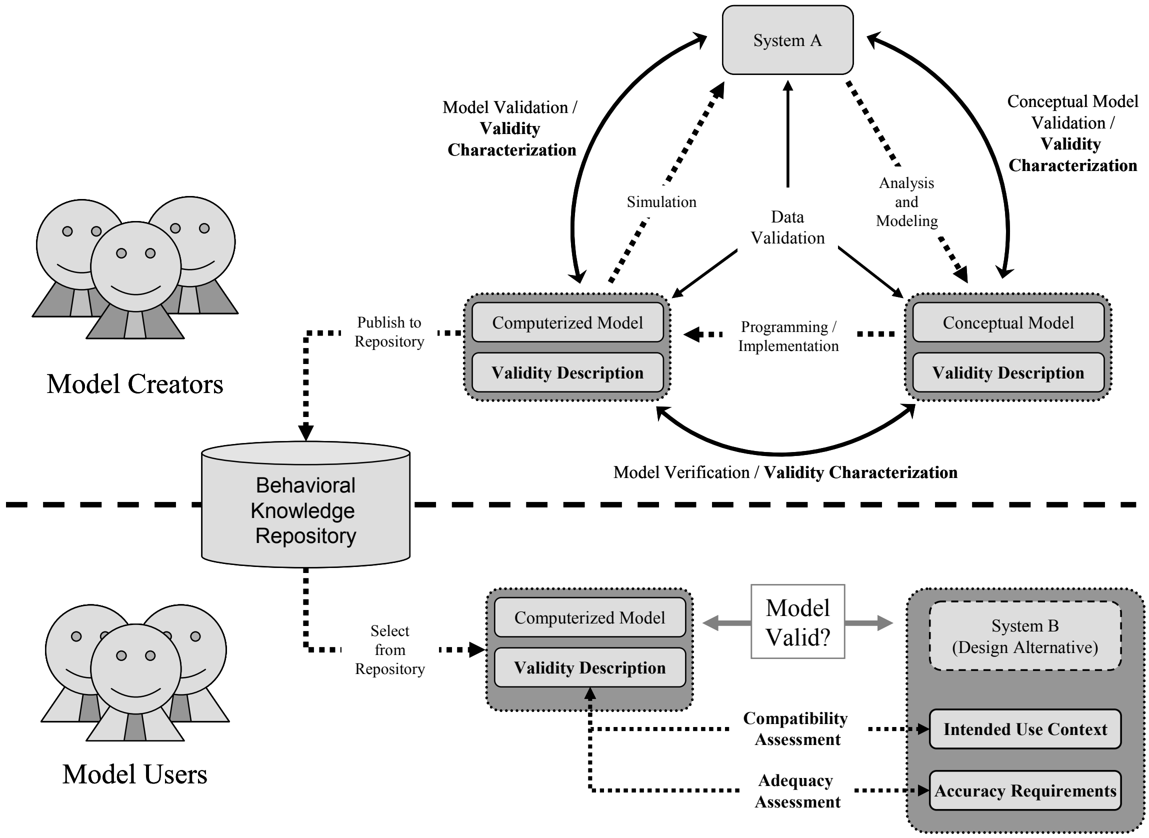

Figure 2.5: A depiction of the model validation problem for reuse scenarios in engineering design with concepts from the hypotheses included.

validity description to an appropriate repository where model users can search for and access them. Model users determine the validity of a model by performing compatibility and adequacy assessment relative to their needs.

2.4 Summary

This chapter contains various perspectives on model validation. Section 2.1 is a fundamental perspective that appeals to philosophical understandings of science and knowledge. Section 2.2 is a practical perspective that reflects fundamental limitations exposed in the philosophy literature and is rooted in an abstract modeling and simulation process. Section 2.3 is a model reuse perspective that exposes the limitations of existing model validation approaches and discusses appropriateness of the hypotheses.

With respect to hypotheses evaluation efforts, this chapter serves to support the following theoretical structural evaluation claims:

- The approaches described in the literature have critical limitations with respect to the problem of behavioral model reuse.
- The hypotheses are appropriate for the problem of behavioral model reuse.
- The hypothesis are compatible with the literature, limitations not withstanding.

The first step in supporting these claims is to report the conclusions of the literature. Section 2.1 is an account of the fundamental limitations common to any model validation approach. It is based on philosophical understandings of science and knowledge. Section 2.2 is an account of the prevailing perspective on model validation. Model validation approaches from the literature are based on a conceptual framework that is described in Section 2.2.2 and depicted in Figure 2.2. This framework is based on a simplified modeling and simulation process and is representative of other frameworks found in the literature.

The first claim is supported in Section 2.3.1. The primary limitation of approaches based on the framework of Figure 2.2 and those similar to it is that it lacks a means to account for validation-relevant knowledge. General model reuse scenarios lead to separations of validation-relevant knowledge from the tasks in which it is needed. One cannot describe the state of validation-relevant knowledge using concepts from the framework of Figure 2.2. The limitations of the framework are summarized in the illustration of Figure 2.4.

Section 0 contains support for the second two claims. The hypotheses are appropriate for the problem of behavioral model reuse because they entail a conceptual framework in which one can account for validation-relevant knowledge properly. They are appropriate with respect to the literature because they preserve the meaning of "validity" and have a meaningful interpretation within existing conceptual frameworks. The concepts identified in the hypotheses and their role in validating reusable behavioral models are depicted in Figure 2.5.

Looking forward, Chapter 3 continues the theoretical structural evaluation of the hypotheses. It includes an elaboration of accuracy quantification and context that further supports their appropriateness. It also includes an explanation of the internal consistency of the framework that completes the theoretical structural evaluation. Chapter 4 contains the bulk of empirical structural and empirical performance evaluations. This is accomplished using example problems to demonstrate the framework. Chapter 5 contains the theoretical performance evaluation and a summary of the overall hypothesis evaluation effort.

CHAPTER 3:

VALIDATING REUSABLE BEHAVIORAL MODELS

The main results of Chapter 2 are that the prevailing conceptual framework for model validation is insufficient for behavioral models reuse scenarios, but that the framework proposed in this thesis is appropriate. However, the discussion in Chapter 2 is relatively abstract. The focus is on establishing that the framework is compatible with the reuse problem. Other theoretical structural evaluation issues, such as substantiating the internal consistency of the hypotheses, are not addressed.

This chapter is an elaboration of the conceptual framework proposed in this thesis and a continuation of the theoretical structural evaluation of the hypotheses. Section 3.1 contains a discussion of *validity descriptions* and their role in the validation of reusable behavioral models. Validity descriptions are comprised of two elements: a description of the total uncertainty—or, *inaccuracy*—in a model and the situations over which this description holds true—or, *context*. Section 3.2 is an explanation of the concept of context. It includes a discussion on how to represent and use knowledge about context. Section 3.3 is an explanation of the concept of inaccuracy. It includes a discussion about different types of uncertainty and uncertainty representations. Section 3.4 is a description of how model creators and model users can perform behavioral model validation in reuse scenarios within the proposed framework. Figure 3.1 contains a summary of objectives of this chapter in relation to the remainder of the thesis.

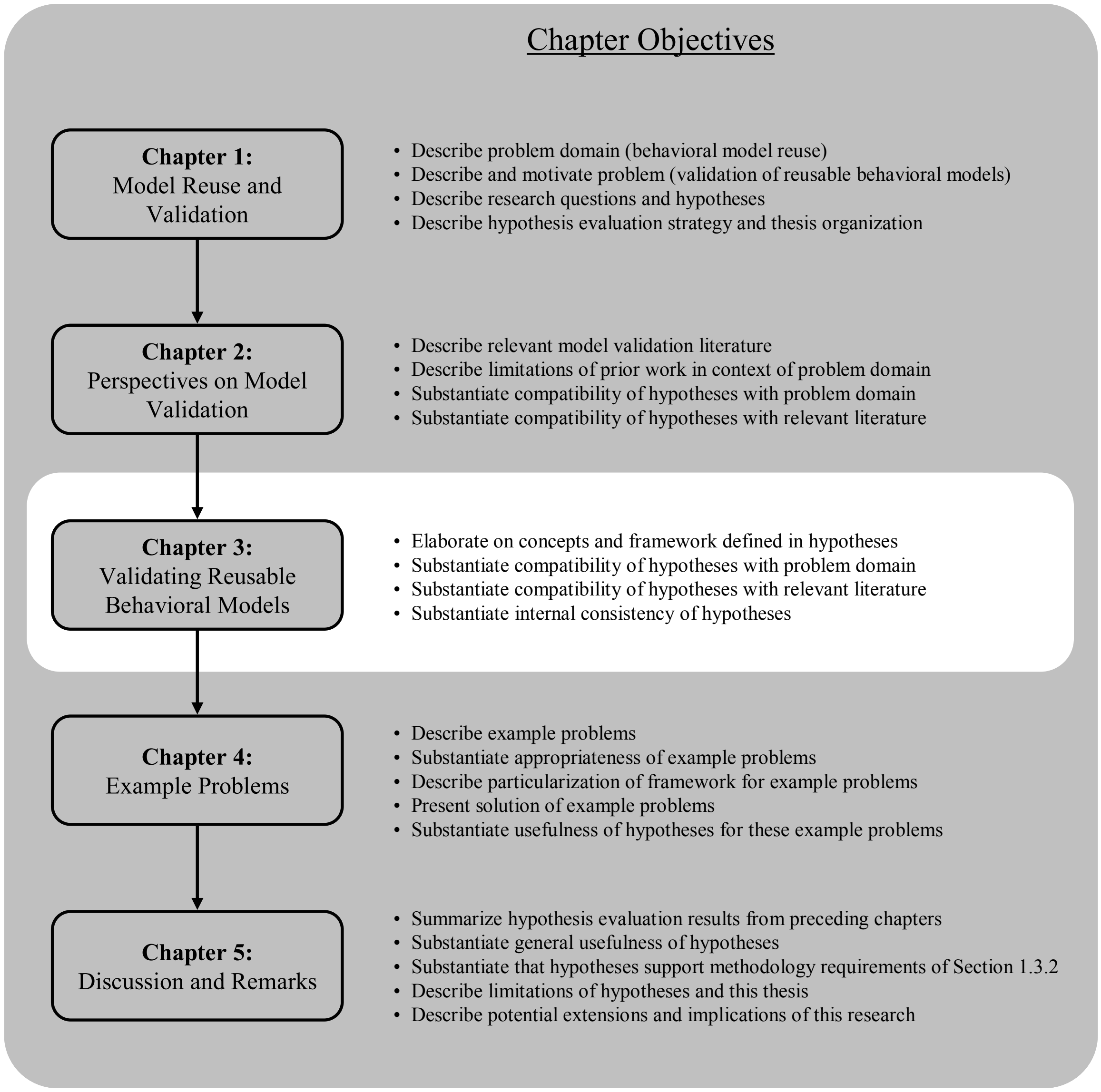

Figure 3.1: Chapter 3 objectives.

3.1 Validity Descriptions and Model Validation

Validity descriptions are central entities in the proposed conceptual framework. A validity description contains the validation-relevant knowledge model users need to establish the validity of a model with respect to their needs. Essentially, it is a conduit for knowledge-transfer between model creators and model users.

A validity description consists of a *context* and *inaccuracy* pair. Context and inaccuracy are formalizations of concepts that are fundamental to model validation. They are defined as follows:

> <u>Context</u>: The limited set of model scenarios over which a statement applies.
>
> <u>Inaccuracy</u>: The total amount of uncertainty in a model or prediction.

These concepts are evident in the definition for behavioral model validation stated in Chapter 1:

> <u>Behavioral Model Validation</u>: The process of determining whether a behavioral model is sufficiently accurate for user needs in the context of intended model scenarios.

The notion of context is explicit in this definition. Inaccuracy is implicit in the statement "sufficiently accurate." Essentially, to determine whether a model is sufficiently accurate one must have knowledge of the total amount of uncertainty present in it. As noted in Chapter 1, most definitions for model validation from the literature include some notion of context and inaccuracy. The meanings of, representations for and use of these concepts are elaborated in Sections 3.2 (context) and 3.3 (inaccuracy).

A validity description is defined as follows.

> Validity Description: A formal statement asserting the inaccuracy of a behavioral model or prediction and a context over which that assertion holds true.

One can interpret a validity description as a guarantee—the creator of the validity description is assuring subsequent users of the associated behavioral model that, within the defined context, the model is no less accurate than the stated inaccuracy. According to the limitations of knowledge (see Section 2.1), this guarantee is not absolute. It represents a model creator's good-faith effort to report the inaccuracy in a particular context. However, conceptualizing this as a guarantee emphasizes the point that validity descriptions are comprehensive, well-supported assertions.

Validity descriptions are not unique. Models may have associated with them different validity descriptions, each of which relates to a different context. Thus, model creators can make several different "guarantees," each of which is true within the stated situations.

Given a validity description, model users can validate the corresponding model without understanding its inner-workings. They only need to understand the formal specification of the validity description and the details of their problem. The use of validity descriptions to this end is apparent in the validation tasks defined by the hypotheses:

> Validity Characterization is the process of creating a validity description for a behavioral model or prediction (Hypothesis 1).

Compatibility Assessment is the process of determining whether the context of a behavioral model or prediction, as stated in its validity description, is compatible with that of the intended use of the model or prediction (step 1 of Hypothesis 2).

Adequacy Assessment is the process of determining whether the inaccuracy of a behavioral model or prediction, as stated in its validity description, is sufficiently small for the intended use assuming it is assessed to be compatible (step 2 of Hypothesis 2).

For a model or prediction to be compatible with an intended use, the context of the use must be more specific than that of the model or prediction. This is to ensure that the inaccuracy assurances given in the validity description for the model or prediction hold true over the extent of the intended use. This condition is discussed in Section 3.2 along with other aspects of context. Whether the inaccuracy of a model or prediction is "sufficiently small" depends on the needs of the intended use. This decision is discussed along with other inaccuracy issues Section 3.3. The process for model validation defined by the hypotheses and its consistency are discussed in Section 3.4.

3.2 Context

This section is an elaboration of the concept of context and a discussion of how one can represent it and reason about it. The way in which one formalizes context is closely related to its meaning. Thus, the two issues are discussed together in Section 3.2.1. Section 3.2.2 is a discussion of how one uses context to draw conclusions about model validity.

3.2.1 Interpretation and Representation of Context

Few statements or rules are universally true. While often left implicit, qualifiers exist that indicate the limitations of information and knowledge. When people communicate, they either presume a common understanding of a domain of discourse or state their assumptions explicitly. Communication is ambiguous when the assumption of common understanding is incorrect. This is particularly dangerous in design, where failure to understand and respect the limitations of knowledge and information can be disastrous. The term *context* refers to the limited domain over which a model or prediction applies.

Several researchers within the artificial intelligence community have discussed the formalization of context for knowledge-based systems (e.g., (Guha, et al. 1992, McCarthy 1993, Akman, et al. 1997); see (Guha, et al. 2003) and (Akman, et al. 1996) for surveys). The general approach they take is to state assumptions about the world as propositions in a logic. Falkenhainer and Forbus take such an approach for describing behavioral model components (Falkenhainer, et al. 1991). The basis for formalizing assumptions comes from the mechanics of mathematical modeling where model creators make simplifications such as assuming a derivative is exactly zero or that a system is completely closed. For example, one can express the position of a particle over time as

$$\mathbf{x}(t) = \int_0^t \int_0^t \mathbf{a}(t)\,dt\,dt + \mathbf{v}_0 t + \mathbf{x}_0,$$

where t is time, $\mathbf{x}_0$ is the initial position vector, $\mathbf{v}_0$ is the initial velocity vector and $\mathbf{a}(t)$ is the acceleration vector as a function of time. Model creators might simplify this model by assuming that acceleration is zero (i.e., constant velocity). In this case, they might report the model using a logical formalization of assumption as

$$\mathbf{x}(t) = \mathbf{v}_0 t + \mathbf{x}_0$$
$$\text{ConstantVelocity} = true$$

where ConstantVelocity is a logical predicate that indicates the constant velocity assumption.

From a mathematical perspective, the use of logical propositions to indicate assumptions is appropriate. In the example above, the semantics of ConstantVelocity are that $\mathbf{a}(t) = {}^{d\mathbf{v}}\!/_{dt} = \mathbf{0}$. That is, each element of the acceleration vector is zero for all time. From a physical perspective, this approach has limitations. Assumptions such as ${}^{d\mathbf{v}}\!/_{dt} = \mathbf{0}$ seldom are satisfied exactly. Even if such an assumption is satisfied exactly, it may be impossible to verify that fact due to limitations in measurements capabilities. Despite this, models that incorporate mathematical assumptions are useful as long as the assumptions correspond "close enough" to reality. In the above example, model inaccuracy is small as long as ${}^{d\mathbf{v}}\!/_{dt} \approx \mathbf{0}$.

The principal limitation of methods that use formal logics to represent the context of a model is that the logical propositions include no indication of how to decide whether an assumption is met "close enough." The person most qualified to make this determination is the model creator. However, this person may not be the same as the user. Model users may lack the domain expertise required to determine how close is "close enough" and, according to the scenarios described in Chapter 1, may be unable to consult model creators.

A set-based approach is more appropriate for representing context. Conceptually, a context defines a set of "world states" within which one has some assurance of correctness or accuracy. There may be no such assurances beyond this region. In principle, a context specifies allowable values of every variable in the "world." In practice, the concept of near-decomposability states that only a handful of variables affect a system (Simon 1996); all others have so little impact on a model's predictions that they can be assumed unbounded. In the simplest of situations, a context set is a hypercube created by bounds on the problem variables. In more complex cases, a context is a region of space defined by functional relationships among the variables and it may include constraints on variables not present in a model. Using a set-based formalism, one might report the model for particle motion under constant velocity as

$$\mathbf{x}(t) = \mathbf{v}_0 t + \mathbf{x}_0$$
$$\left\| \frac{d\mathbf{v}}{dt} \right\|_2 \leq \mathbf{a}_{\max}$$

where $\|\cdot\|_2$ is the Euclidean norm ($\|\mathbf{u}\|_2 = \left(\mathbf{u}^T\mathbf{u}\right)^{1/2}$) and $\mathbf{a}_{\max}$ is a small positive upper limit on magnitude of the acceleration vector. This formulation of context is unambiguously interpretable by model users.

3.2.2 The Role of Context in Model Validation

A context forms the basis of a validity description and is therefore central to the validation of reusable behavioral models. For decision problems, each behavioral prediction contributing to a decision must satisfy its own contextual obligations. These obligations relate to which aspect of behavior—or which *behavioral attribute*—a decision maker (i.e., model user) wants predicted. Decision makers typically require predictions about different behavioral attributes of a system, and each behavioral attribute can have a different context. For example, a decision maker might require one prediction about structural stress under steady-state conditions and another about the probability of failure under specific dynamic conditions. Context requirements for a particular behavioral attribute in a particular decision problem are referred to as a *behavioral attribute context*. A decision maker performs compatibility assessment for a prediction by comparing its context to that of the corresponding behavioral attribute. A decision maker can use a prediction only if it applies over the entire behavioral attribute context. Otherwise, there will be portions of the behavioral attribute context in which the prediction cannot necessarily be trusted. Decision makers take decisions in this circumstance at their own risk; to do so would be like making decisions about a supersonic aircraft based upon subsonic performance predictions.

In general, one can rationally execute a decision if and only if each prediction is of the same or broader context than its corresponding behavioral attribute—that is, they must subsume the behavioral attribute contexts. Figure 3.2 contains conceptual depictions of two possible decision making scenarios, each with a behavioral attribute context and the context of a corresponding prediction. In the situation depicted in Figure 3.2(a), one can make a rational decision because the context of the prediction information subsumes the behavioral attribute context. That is, they are *context-compatible*. One cannot make a rational decision in the situation depicted in Figure 3.2(b). Here, the prediction and the behavioral attribute are not context-compatible. All is not lost if the context requirements for a decision cannot be met at first. It is often possible to expand the context of a prediction if one is willing to trade a degree of accuracy for it (accuracy and the context-accuracy relationship are discussed in the next section).

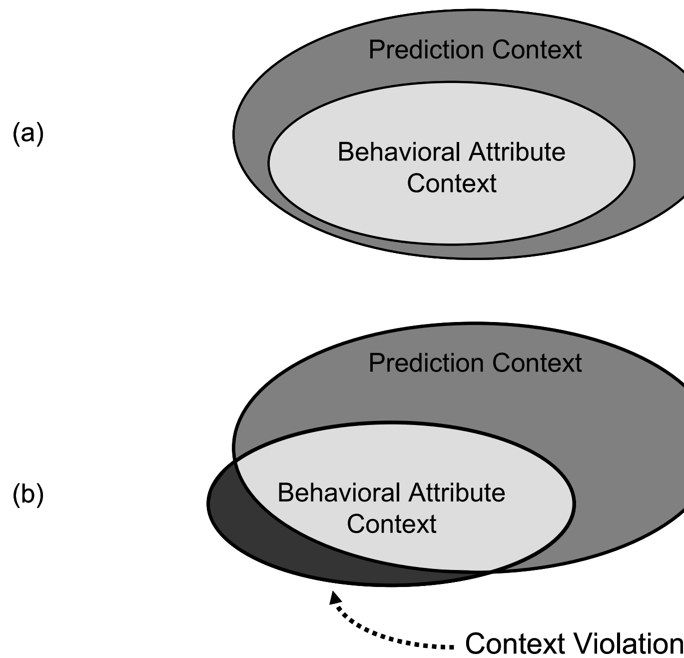

Figure 3.2: Contexts in a design decision: (a) the behavioral attribute context is subsumed by that of the corresponding prediction; (b) the behavioral attribute context is not subsumed by the context of the corresponding prediction.

A simulation experiment is comprised of a model and the inputs and parameters for the model. For a design problem, parameters specialize a behavioral model to a particular design alternative (i.e., they specify physical dimensions or other quantities that remain constant throughout the simulation) and inputs represent external stimuli. Each element of a simulation experiment is associated with a particular context and the context of a prediction made by the simulation is the intersection of these contexts.

A simulation experiment is comprised of a model and the inputs and parameters for the model. For a design problem, parameters specialize a behavioral model to a particular design alternative (i.e., they specify physical dimensions or other quantities that

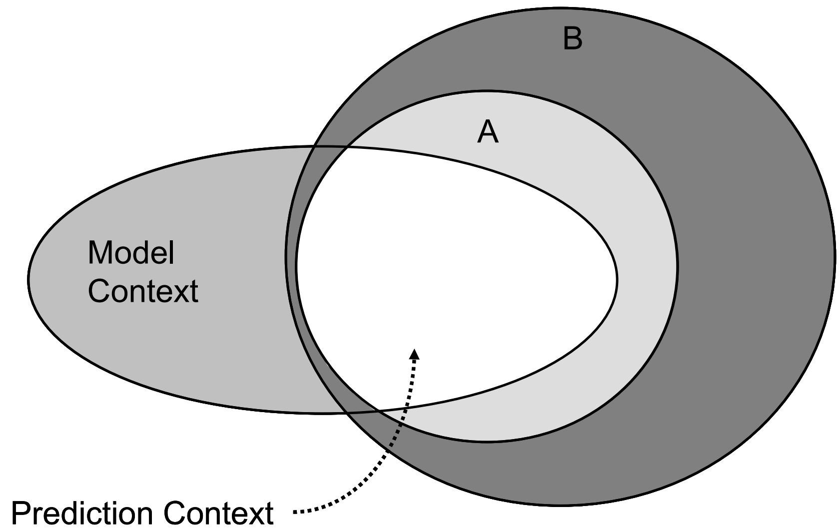

Figure 3.3: A depiction of the relationship between the context of a prediction and those of the model and information (model parameters or inputs), A and B, from which it is formed.

remain constant throughout the simulation) and inputs represent external stimuli. Each element of a simulation experiment is associated with a particular context and the context of a prediction made by the simulation is the intersection of these contexts. Figure 3.3 contains a conceptual depiction of the relationship of a prediction context to the contexts in a simulation experiment. Intuitively, a prediction cannot "know" more than the elements from which it was formed. For example, one cannot generally make valid predictions about turbulent fluid flow based solely on a laminar flow model. Mathematically, one can state the general relationship as

$$C_P = C_M \cap \left(\bigcap_{j=1}^{n} C_j \right), \tag{3.1}$$

where C_P is the context of the prediction, C_M is the model context and C_j is the context of the j[th] input or parameter to the model. This means that the context of a prediction is never more general than the least general context from which it is formed.

One can assess the compatibility of a model on two levels. First, one can ask whether a model is compatible with a given simulation experiment. To answer this question, one compares the context of a model to those of the parameters and inputs of an experiment. A model is compatible with the other elements of the experiment if the intersection of these contexts—i.e., the context of the resulting prediction—is not the empty set.

More commonly in design, one performs a simulation experiment to predict a specific behavioral attribute for use in a decision. In this case, one performs compatibility assessment for a model relative to whether a resulting prediction is context-compatible with the behavioral attribute. Given inputs and parameters for a model and a desired behavioral attribute, the use of a behavioral model is valid if the prediction yielded by the simulation experiment is context-compatible with the behavioral attribute. This combines the concepts illustrated in Figure 3.2 and Figure 3.3. Combining the relationship in Equation (3.1) with the notion of context compatibility yields

$$\mathrm{ContextCompatible}\left(C_{BA}, E\right) \Leftrightarrow C_{BA} \subseteq \left(C_M \cap \left(\bigcap_{j=1}^{n} C_j\right)\right), \qquad (3.2)$$

where $\mathrm{ContextCompatible}(\cdot,\cdot)$ is a logical predicate, C_{BA} is the behavioral attribute context and E is a simulation experiment definition that specified model, input and parameter contexts. With respect to validating the use of a behavioral model, the condition $C_{BA} \subseteq C_M$ is necessary for Equation (3.2) to hold.

In general, one may be unable to complete compatibility assessment before performing a simulation. This is the case when model inputs are time-dependent and the full input trajectory is unknown prior to performing simulation. For instance, some simulations consist of multiple models. The context compatibility of one model can depend on the output of another. One must ensure that context compatibility is maintained throughout the course of a simulation.

3.3 Inaccuracy

This section is an elaboration of the concept of inaccuracy and a discussion of how one can represent it and reason about it. Section 3.3.1 is a description of the interpretation of inaccuracy and different formalizations for representing it. Section 3.3.2 is a discussion of how one uses inaccuracy during model validation.

3.3.1 Interpretation and Representation of Inaccuracy

Behavioral models are approximations of real systems. As such, models and any predictions derived from them can differ from reality. Although it is impossible to know system behavior with certainty, it is possible for one to gauge the degree of uncertainty in their knowledge. *Inaccuracy* refers to the total amount of uncertainty in a prediction or model.

One source of uncertainty in modeling and simulation is natural variability. Uncertainty due to random variability is referred to as *aleatory uncertainty* (Parry 1996). Authors use different terminology to refer to this concept, including variability, stochastic uncertainty, objective uncertainty (Ferson, et al. 1996) and irreducible uncertainty. Examples of phenomena that involve or exhibit aleatory uncertainty include machining error, annealing, errors in communication systems, many measurement errors and radioactive decay.

Another source of uncertainty is incomplete knowledge. This type of uncertainty is referred to as *epistemic uncertainty* (Parry 1996). It sometimes is called reducible uncertainty, subjective uncertainty (Ferson, et al. 1996) or, in the context of decision problems, imprecision (Antonsson, et al. 1995). Epistemic uncertainty often results from ignorance or modeling decisions, such as selecting one model over another or choosing to make particular approximations and simplifications. Oberkampf and coauthors distinguish between epistemic uncertainty and error, which they describe as resulting from deliberate simplifications or inadvertent mistakes in modeling (Oberkampf, et al. 2002). However, error is a type of knowledge deficiency and is therefore better viewed as a subclass of epistemic uncertainty.

Because aleatory uncertainty is a result of randomness, one can represent it using classical probability theory. Probability theory allows one to express the possible outcomes of a random event and their relative likelihoods. This corresponds to the extent of knowledge one can have about an aleatory uncertainty: one cannot know a priori the result of a particular trial, but one can know the aggregate results of a population of trials.

One cannot use classical probability theory to represent epistemic uncertainty. The underlying cause of epistemic uncertainty is incomplete knowledge. An assumption underlying classical probability theory is that one has complete knowledge about a random trial—the set of possible results and the relative likelihood of the results. By lacking some amount of knowledge, one cannot complete the probabilistic representation without incorporating assumptions. For example, assume an engineer wishes to formalize the inaccuracy in a parameter and knows only that the parameter value is between 5 and 10. With no knowledge of how likely any of the values are, the engineer might be tempted to represent the inaccuracy using a uniform distribution ranging from 5 to 10. However, strictly speaking, this is incorrect. The engineer has no basis for assuming that, say, 5.3 is equally as likely as 6.9. A more appropriate inaccuracy representation would be an interval ranging from 5 to 10. In the case of the simple example, this representation incorporates no unsubstantiated assumptions.

In general, formal approaches for representing and making decisions under epistemic or combined epistemic-aleatory uncertainty are a topic of ongoing research. Investigators have explored several alternatives to classical probability theory, including possibility theory (Dubois 1988), fuzzy set theory (Zadeh 1965), Dempster-Shafer theory (Schafer 1976, Yager, et al. 1994), probability-bounds analysis (Ferson 2000), interval analysis (Ferson, et al. 1996) and set theory ((Ben-Haim 2001) as part of information gap decision theory).

3.3.2 The Role of Inaccuracy in Model Validation

The notion of inaccuracy is important to designers who use models because they must contend with both aleatory and epistemic uncertainty. Aleatory uncertainty is particularly important when considering the impacts of manufacturing variations (e.g., random deviations in part sizes) and variations in interactions with external systems (e.g., random deviations in loading conditions). One source of epistemic uncertainty is the act of modeling. Because all models are approximations of reality, they have epistemic uncertainty and, by virtue of being computed from a model, all predictions have epistemic uncertainty as well. Another source of epistemic uncertainty is the incompleteness of a design specification (i.e., there is a lack of knowledge about what the final design will be). This source of uncertainty manifests itself in behavioral models since a model cannot "know" more about a design than is present in its specification. Also note that there can be epistemic uncertainty about an aleatory uncertainty. For example, one may not know the precise mean of a probability distribution.

The purpose of a validity description is to provide model users with assurances about the inaccuracy of a behavioral model or prediction over a well-defined set of situations. The inaccuracy assurances are then used during adequacy assessment. In general, users have no way of knowing if the inaccuracy of a model or prediction is actually larger than what is reported. Because of this, model creators must ensure that their characterizations of inaccuracy do not understate the actual inaccuracy. This suggests that inaccuracy characterizations should be conservative. For interval-based inaccuracy representations, a conservative characterization is wider than one that is non-conservative. For probabilistic representations, one might be conservative by stating a larger variance than for a non-conservative characterization. Overly conservative characterizations are undesirable because they artificially limit the usefulness of an item. However, it is better for model creators to error on the side of conservativeness than to take a chance that understating the inaccuracy will not matter.

There are two main alternatives for representing inaccuracy due to combined aleatory-epistemic uncertainty. One alternative is to represent inaccuracy using an interval or, more generally, a set-based approach. Under a pure set-based approach one seeks a bound on the inaccuracy and does not consider distribution information about the aleatory information. Ideally, one finds the least upper bound, or *supremum*, of the inaccuracy set. However, any upper bound will do. This approach is conservative, but fails to utilize all of the available information. A second alternative is to use an approach that is capable of representing the aggregate aleatory-epistemic uncertainty. One such approach is called *probability bounds analysis* (Tucker, et al. 2003). This approach combines interval analysis with probabilistic methods. Using this approach, one can represent inaccuracy as a pair of cumulative distribution functions that bound the set of possible distribution functions. In this way, one can establish a conservative bound on the inaccuracy while still incorporating available probabilistic information about aleatory uncertainty.

Users assess the adequacy of a model by comparing their accuracy needs to the inaccuracy assertion in the validity description associated with the model. A model is adequate if its inaccuracy is "less than" the maximum inaccuracy tolerable by the user. One's interpretation of "less than" can depend on the inaccuracy representation one adopts. For example, one may adopt a strict subsumption-based interpretation for interval-based representations but a confidence-level based interpretation for probabilistic representations.

For most models and predictions, inaccuracy depends upon the context in question. For instance, a linear deflection model for a beam may be very accurate when the displacement is less than some upper bound, but inaccurate otherwise. In general, inaccuracy never decreases—and likely increases—as the context expands. This results in an important tradeoff for model creators: too narrow a context can yield a very accurate model that is seldom useful, while too broad a context can result in a model too inaccurate to be useful.

3.4 Performing Model Validation within the Proposed Framework

This section is an elaboration of model validation process implied by the hypotheses proposed in this thesis. Section 3.4.1 is a detailed description of the process flow using the concepts introduced in this thesis. Section 3.4.2 is an explanation of how this process is internally consistent and preserves the semantics of validity that are accepted in the literature.

3.4.1 Process Flow

The notions of context and inaccuracy are useful beyond the conceptual framework proposed in this thesis. This is evidenced by their notional alignment with definitions for model validation (see Section3.1). One could use them when performing model validation within a framework such as that depicted in Figure 2.2 However, the significance of the concepts is that they also are appropriate for the more general problem of validating reusable behavioral models. Together, context and inaccuracy form the basis for a fundamental unit of validation-relevant knowledge—a validity description.

Figure 3.4 is a flow chart for a general model validation process that is appropriate for model reuse scenarios. The figure is an elaboration of Figure 1.3 using the concepts and relationships from the framework proposed in this thesis. It includes explicit indication of the decision points that model users encounter. It also includes steps from the overall modeling and simulation process for reusable behavioral models (see Figure 2.3). Note that as in Figure 2.3, the flow depicted consists of two processes—model creation and model reuse—that operate independently. Knowledge exchange

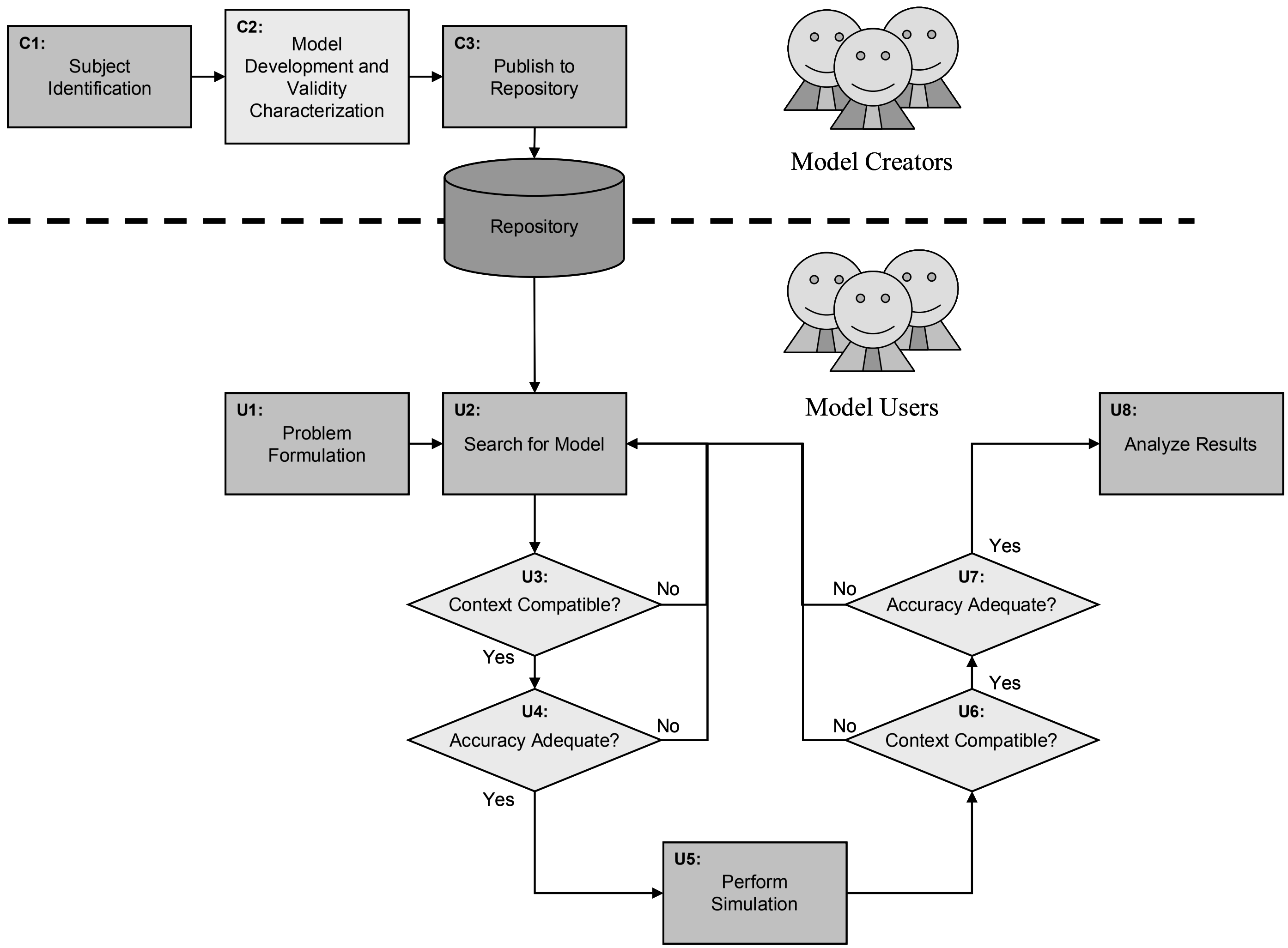

Figure 3.4: Flow chart for model validation process based on the proposed conceptual framework.

between the processes occurs via the repository contents (behavioral models with corresponding validity descriptions).

Model Creation

In step C1, model creators identify a subject matter to be modeled. The output of this step is a system and set of scenarios in which the model should reflect the behavior of the system. In practice, model creators may revisit this step as they learn more about the system.

Step C2 involves model development and validity characterization. Creators develop a model that represents a specified system over the specified set of scenarios. In accordance with hypothesis H1, creators also develop a validity description of the model that includes an assertion about its inaccuracy over a well-defined context. In some instances, model development and validity characterization may be sequential steps in the creation process. In general, they can proceed concurrently.

Step C3 involves publishing a completed model and its validity description to a suitable repository. In this step, creators may formalize meta-information to make it easier for model users to locate models that correspond to their needs. This information might include a description of the system being modeled, the interface to the model (e.g., its inputs, outputs and parameters) and the model implementation (e.g., Matlab, Java, C).

Model Use

In step U1, model users formulate a problem to be solved using modeling and simulation. They must identify the system to be modeled and the scenarios in which it must be modeled. This step is analogous to C1 in the creation process.

In step U2, users search a repository for candidate models. In this search, they identify models that are appropriate for use if they prove to be valid. Such models represent the system of interest to the users and satisfy any implementation constraints. The models must have appropriate interfaces and be executable by the users.

Steps U3 and U4 are preliminary executions of the validation steps of hypothesis H2—compatibility assessment and adequacy assessment. Steps U6 and U7 are final executions of these steps. A two-pass approach is necessary when one cannot determine all of the conditions of a simulation without actually performing it. For instance, a dynamic simulation might be within the context of a model at the outset, but stray beyond it as the simulation progresses. In such situations, the first pass (steps U3 and U4) eliminate obviously invalid models and the second pass (steps U6 and U7) serve to confirm that the chosen model is valid for the entire simulation (performed in step U5).

Steps U3 and U6 correspond to the first validation step of hypothesis H2—compatibility assessment. In step U3, users determine whether the context stated in the validity description of a model subsumes that of the problem identified in step U1. The context of the problem reflects the initial conditions of the simulation, but not necessarily conditions throughout the entire simulation. In step U6, one again performs compatibility assessment, but this time with respect to the context encountered during the simulation. Both compatibility assessment steps involve the context-compatibility test identified in Equation (3.2) of Section 3.2.2. If either test fails, the process returns to step U2 for the selection of a new model.

Steps U4 and U7 correspond to the second validation step of hypothesis H2—adequacy assessment. In step U4, users determine whether a model is sufficiently accurate for their needs. Sometimes, one will be unable to make a definitive judgment about adequacy prior to conducting a simulation. For instance, this can happen when one is comparing alternatives and must know the nominal performance of each before determining what level of inaccuracy is tolerable. In step U7, one performs a final adequacy assessment if the simulation results matter in the adequacy decision. If a model fails either in step U4 or U7, one discards it and returns to step U2 to select another model.

Step U5 involves the use of a model that passes the compatibility and adequacy assessment steps. Step U8 consists of final results analysis, excluding the final validation checks (steps U6 and U7). In this step, users make determinations about the validity of the overall simulation experiment (as opposed to that of the model; see Section 1.2.3).

3.4.2 Appropriateness and Consistency

The process of Figure 3.4 results from a combination of the proposed framework and the modeling and simulation process for model reuse depicted in Figure 2.3. The model validation steps defined in the framework fit cleanly within the model creation and model use processes. This is necessary for the proposed framework to be consistent with behavioral model reuse scenarios.

The framework leads to a sequence of steps that are consistent with one another. This is evidenced by the flow diagram of Figure 3.4. At the inception of each step, the requisite knowledge and information is available so that one can achieve the objectives of step.

- Validity characterization (part of step C2) requires a well-defined system relative to which one can characterize a model. This is output from step C1.

- Preliminary compatibility and adequacy assessment (steps U3 and U4, respectively) require a well-defined problem and a model with an associated validity description. The former is a result of step U1 and the latter is a result of model development (steps C1-3).
- Final compatibility and adequacy assessment (steps U6 and U7, respectively) require a well-defined problem, a model with an associated validity description and the results of a simulation experiment. A problem description is a result of step U1. A model and its validity description result from steps C1-3. Simulation results are an output of step U5.

Internal inconsistencies could result in circular dependencies or missing information and knowledge. These problems are not present.

Strictly speaking, one could omit steps U3 and U4 from the flow diagram. By performing steps U6 and U7, one could reject the results of an invalid model regardless of whether one performs preliminary assessments. Thus, the functions defined in hypothesis H2 are achieved. However, steps U3 and U4 are important from an efficiency standpoint. One should rule out obviously invalid models prior to using them. Although a two-pass strategy is not spelled out in the hypothesis explicitly, it is compatible with the hypothesis and the flow chart of Figure 3.4 is appropriate to serve as evidence of consistency.

Given the proposed framework leads to an internally consistent model validation process, it is reasonable to question whether it is appropriate. That is, will following the specified steps lead a model user to establish the validity of a behavioral model? One can find an answer to this by examining the definition for behavioral model validation. According to the definition from Chapter 1, one can consider a model to be valid if it is "sufficiently accurate" within the "context of intended model scenarios." One can observe similar language in the alternate definitions of Table 1.2. An appropriate conceptual framework for validating reusable behavioral models includes the concepts and processes necessary to make this determination. Context and inaccuracy, which comprise a validity description, form a conceptual basis for this determination. A model that is context compatible with a particular use satisfies the statement "in the context of intended model scenarios." To say that a model is "sufficiently accurate" is equivalent to saying that its inaccuracy is less than the maximum allowable inaccuracy for a given problem. Thus, a model that passes the compatibility and adequacy assessment steps is valid in the sense of the accepted meaning of validity.

In addition to being compatible with the prevailing meaning of validity and being internally consistent, the proposed framework also reflects the requirements for validation of reusable behavioral models as stated in Table 1.3. The requirements and how they are met are summarized in Table 3.1.

The first requirement is that time spent performing validation activities at the time of model use be made small. The framework addresses this by including only knowledge use operations (e.g., the assessment steps) in the model use process. Validity characterization, the task of developing validation-relevant knowledge about a model, is performed in an independent model creation process. One can observe this arrangement of tasks in the two processes in Figure 3.4. This arrangement keeps validation activity time in the use process small relative to the total time invested in validation activities across both processes because knowledge creation typically is much more time consuming than knowledge use.

Table 3.1: Requirements for validation schemes involving model reuse and how they are met in the proposed conceptual framework.

No.	Requirement	How Met?
1	The time spent performing validation activities at the point of model use must be made small.	Validity characterization performed in separate model creation process.
2	Validation-relevant knowledge must be represented explicitly and associated with behavioral models.	Validity descriptions.
3	Validation-relevant knowledge must be described in terms of concepts that have well-defined semantics that are independent of any particular person, group or project.	Context and inaccuracy in a validity description.
4	Validation-relevant knowledge must be expressed in a mathematically formal manner.	Context and inaccuracy in a validity description.

The final three requirements are addressed by use of validity descriptions. The second requirement is that one must represent validation-relevant knowledge explicitly and associate it with the corresponding behavioral model. The third requirement is that one must describe validation-relevant knowledge using concepts that are semantically well-defined and that are independent of any person, group or project. The final requirement is that one expresses validation-relevant knowledge in a mathematically formal manner. Validity descriptions are an explicit representation of the validation-relevant knowledge identified in Section 1.2.2. This knowledge is developed and formalized during validity characterization. The notions of context and inaccuracy have well-defined meanings with respect to the model validation problem that are invariant across different modeling and simulation applications. Furthermore, one specifies context and inaccuracy in a mathematically formal manner. The conceptual framework does not contain strict constraints on what formalism one must use because appropriate formalisms, particularly for inaccuracy, are an open area of research. However, intentions are that one will use an appropriate formalism.

3.5 Summary

This chapter is an elaboration of the proposed conceptual framework for behavioral model validation and a continuation of the theoretical structural evaluation of the hypotheses. Section 3.1 is a discussion of validity descriptions and how one uses them to represent validation-relevant knowledge about a behavioral model. Sections 3.2 and 3.3 are explanations of context and inaccuracy, respectively. These are the elements of a validity description. Section 3.4 contains a process flow chart for the validation activities in the model creation and model use processes (Figure 3.4). It also contains a discussion about the appropriateness and consistency of the proposed conceptual framework.

With respect to hypotheses evaluation efforts, this chapter serves to support the following theoretical structural evaluation claims:

- The hypotheses are internally consistent.
- The hypotheses are compatible with the problem domain of validating reusable behavioral models.
- The hypotheses are compatible with the relevant literature and existing interpretations of model validity.

Support for each of these claims appears throughout the chapter. Section 3.4 contains a more focused discussion. Support for internal consistency results from the flow diagram of Figure 3.4 and discussion of the process. Support for compatibility with the problem domain includes two primary considerations. First, the flow diagram of Figure 3.4 is essentially a more detailed version of the model reuse process of Figure 2.3. This means that the process flow is compatible with that of model reuse in engineering design. Second, the proposed hypotheses address the requirements for behavioral model validation schemes in reuse scenarios. These requirements are established in Section 1.3.2. The way in which the hypotheses address them is described in Section 3.4.2 and summarized in Table 3.1. Support for compatibility with the relevant literature follows from an examination of accepted definitions of model validation. Context and inaccuracy are formalized representations of concepts present or implied in these definitions.

The next chapter contains the definition and solution of two example problems. The objectives behind these example problems are to demonstrate the concepts in action and to provide support for hypothesis evaluation. Specifically, Chapter 4 contains empirical structural and empirical performance evaluations of the hypotheses. The final chapter, Chapter 5, contains the theoretical performance evaluation and a summary of the overall hypothesis evaluation effort.

CHAPTER 4:

EXAMPLE PROBLEMS

This chapter contains the definition and solution of two example problems. These serve to illustrate the proposed conceptual framework and to support hypothesis evaluation activities. Specifically, the contents of this chapter support empirical structural and empirical performance evaluations. Figure 4.1 is a summary of the chapter objectives and their role in the thesis.

Section 0 contains a preliminary discussion about the example problems. It includes a discussion of why the examples are appropriate for supporting hypothesis evaluation. It also includes a description of how the abstract conceptual framework is particularized for these examples. Sections 4.2 and 4.3 contain the example problems. The first involves validity characterization of a formulation of Newton's second law of motion and compatibility and adequacy assessment of that model for a particular use. The second involves validity characterization of a more complex model, a beam in axial tension. These sections comprise the empirical performance evaluation of the hypotheses. Both include remarks about the results and their implications about the hypotheses. In particular, the remarks include a discussion of the usefulness of the conceptual framework for solving the example problems. Chapter 5 contains discussion about the extension of the conceptual framework to other examples.

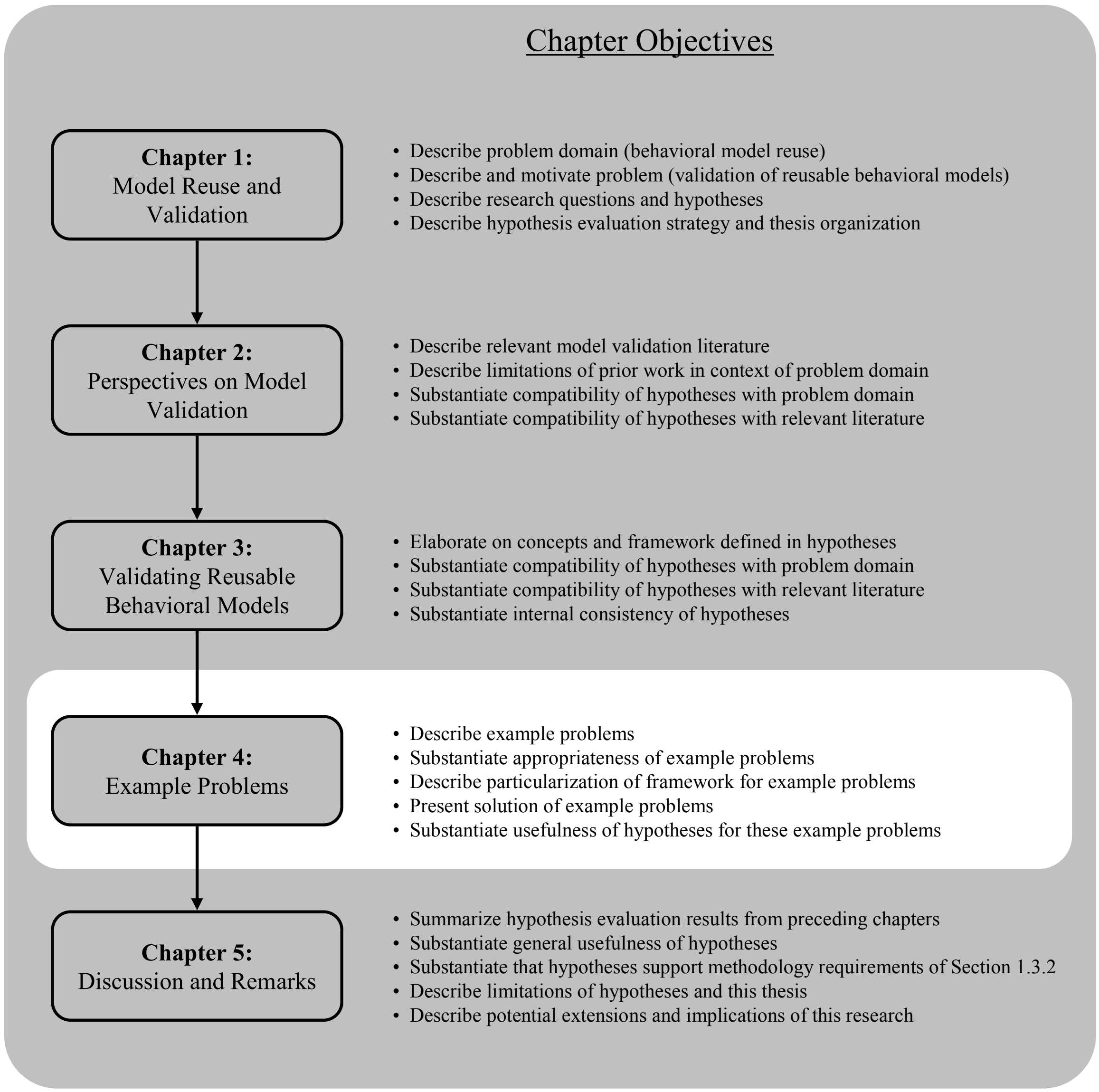

Figure 4.1: Chapter 4 objectives.

4.1 Preliminary Comments on the Example Problems

This section contains remarks on the example problems and their solution. Section 4.1.1 is an explanation of the appropriateness of the example problems. In terms of the hypothesis evaluation activities outlined in Section 1.4.2, this constitutes empirical structural evaluation. Section 4.1.2 is a discussion of the particular methods used in the examples. These methods are consistent with the conceptual framework defined in the hypotheses but are not themselves part of the framework. Specific methods are necessary to implement the abstract portions of the framework.

4.1.1 Overview and Appropriateness

Individually, an example is appropriate if it exhibits one or more of the problem domain characteristics. As a set, the example problems must span the problem domain characteristics. This allows one to determine whether, relative to these examples, the hypotheses are appropriate over the entire problem domain. In this thesis, the problem domain is the validation of reusable behavioral models. The salient characteristics of this domain are discussed at various points in this thesis. They include:

- Model creators have no specific knowledge of how their models might be reused.
- Model users have no specific validation-relevant knowledge about a model other than what is included in a validity description.
- Model users may not have access to empirical data relative to which they can perform model validation.

Each of the examples in this chapter reflect one or more of these problem domain characteristics and together they span the set of characteristics.

The example of Section 4.2 includes complete validation activities from both the model creation and model use processes (see Figure 4.2). As such, it reflects all of the above domain characteristics. Model creators perform validity characterization (step C2) on a formulation of Newton's second law of motion. No knowledge of subsequent model use is present during characterization. Model users then perform validation relative to a particular use of the model using only the knowledge representing in the corresponding validity description. This consists of compatibility assessment (step U3) and adequacy assessment (step U4). The secondary assessment steps (steps U6 and U7) are not necessary because the entire model use context is known prior to evaluating the model. The outputs steps C1 and U1 are given as part of the example problem definition. The steps therefore are not carried out in Section 4.2. The steps in Figure 4.2 relating to a model repository (steps C3 and U2) are not carried out explicitly since we are dealing with only one model. Omission of these steps does not detract from the results of the examples, since the objective is to examine the suitability of the conceptual framework for dealing with validation relevant knowledge.

Although the exercise is not performed by two independent groups of people, the independence of the two processes is evident from the flows of information and knowledge in the example. The model used for this example is simple relative to some engineering models. This is beneficial from an expositional standpoint because it prevents the details specific to the example from occluding those of the conceptual framework.

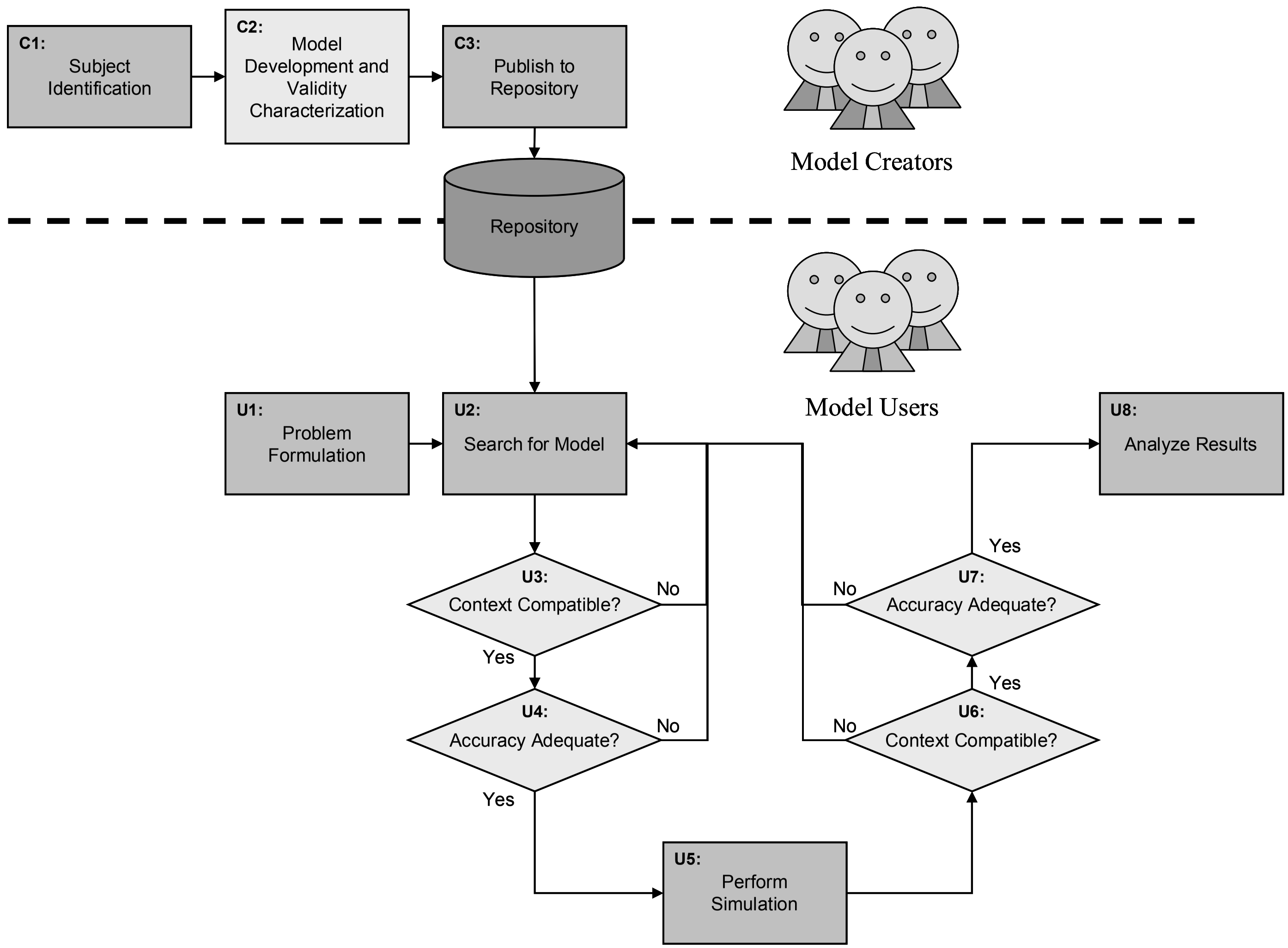

Figure 4.2: Model validation process from Figure 3.4.

The example of Section 4.3 reflects only the first of the above domain characteristics, but it is a deeper treatment of validity characterization (step C2 in Figure 4.2). It involves the development and characterization of a model for deformation in an axially loaded beam. The reason for focusing on validity characterization in lieu of compatibility and adequacy assessment is that, given an appropriate validity description, the assessment steps are largely the same as those encountered in the first example. Because the assessment steps depend on an well-formed validity description, an exploration of validity characterization is an important step in evaluating the conceptual framework.

4.1.2 Adopted Methods and Representations

Because the hypotheses advanced in this thesis are conceptual in nature, it is necessary to assume concrete methods in order to solve the example problems. For instance, H1 states that model creators can develop validity descriptions of their models but includes no indication of how one should represent the knowledge contained in a validity description. Chapter 3 contains a discussion about representation formalisms, but the discussion is inconclusive. Without settling the representation issue, it is impossible to identify concrete methods for performing compatibility and adequacy assessment.

The methods adopted in this chapter are consistent with the conceptual framework defined in the hypotheses, but they are not the only possible selections. Thus, they allow one to evaluate the hypotheses on a feasibility basis. The success of these methods on these examples serves as a proof-of-concept for the framework.

Validity characterization in the example problems proceeds in a two-step process:

1. Assume a context for the model
2. Determine the inaccuracy of the model over that context

A restricted set-based representation is adopted for context. Context sets are defined by bounds on vector norms for variable values. For a given variable, the context includes all values for which the magnitude is less than the stated bounds. For vector variables, a Euclidean norm is used to measure magnitude. This is a special case of the set-based approach discussed in Chapter 3

Determination of context bounds is at the discretion of model creators. The rationale underlying such decisions is beyond the scope of this thesis. Of relevance to this work is that model creators determine the bounds without direct knowledge of future model uses.

Inaccuracy is represented using a set-based approach. This reflects an epistemic uncertainty in which one has no knowledge of the relative likelihood of set members. An inaccuracy parameter is defined for a model relative to other knowledge taken to be "ground truth." A least upper bound for this parameter over the context set is found.

Both example problems include a problem statement for model use. This statement defines the "givens" and "finds" of a simulation problem. It also includes definitions for the physical scenarios of interest (i.e., context of interest) and the maximum level of inaccuracy that is allowable.

Compatibility assessment involves a comparison of interval bounds for model context and simulation problem context. This is a direct evaluation of Equation (3.2) for the special case of an interval-based context representation.

Adequacy assessment involves a comparison of model inaccuracy to the maximum allowable inaccuracy specified in the problem statement. For this comparison, it is necessary to determine the impact of model inaccuracy on the variable to be predicted. All other variables are assumed to be perfectly accurate and the model inaccuracy is propagated to the prediction. The magnitude of prediction inaccuracy is compared to the inaccuracy requirements of the example problem to decide whether the model is adequate.

4.2 Example Using a Formulation of Newton's Second Law of Motion

This section contains an example of how one can perform model validation for a reusable behavioral model according to the hypotheses proposed in this thesis. The example involves a model of low complexity in order to focus on the details of the model validation framework. The example is comprised of two steps that correspond to the validation activities performed within the model development and model use processes as depicted in Figure 3.4. Section 4.2.1 is an account of the validity characterization process for the model. This corresponds to step C2 in the flow chart. Section 4.2.2 is an account of the compatibility and adequacy assessment steps for a particular simulation problem. These correspond to steps U3 and U4 in the flow chart. Section 4.2.3 contains remarks about how this example supports the appropriateness of the proposed hypotheses.

4.2.1 Validity Characterization

A common formulation of Newton's second law of motion is

$$\mathbf{F}(t) = m\mathbf{a}(t) \tag{4.1}$$

where $\mathbf{F}(t)$ is the net force vector on a particular as a function of time, m is the particle mass and $\mathbf{a}(t)$ is the particle acceleration vector as a function of time. One may recognize that this is a simplification of the more general relationship

$$\mathbf{F}(t) = \frac{d}{dt}\left(m(t)\mathbf{v}(t)\right) \tag{4.2}$$

where $m(t)$ is the time-varying particle mass and $\mathbf{v}(t)$ is the time-varying velocity vector. Expanding the derivative, one has

$$\mathbf{F}(t) = m(t)\dot{\mathbf{v}}(t) + \dot{m}(t)\mathbf{v}(t)$$

where $\dot{m}$ and $\dot{\mathbf{v}}$ are the time-derivatives of mass and velocity, respectively. This relationship is a more accurate reflection of reality than the one given in Equation (4.1). Because the difference between the two models is the term $\dot{m}(t)\mathbf{v}(t)$, the inaccuracy of Equation (4.1) depends on the particle velocity and time-derivative of particle mass.

When performing validity characterization, one's objective is to determine the inaccuracy of a model over some fixed and well-defined context. In the current example, one must select a context that bounds particle velocity and time-derivative of mass. Without doing so, the inaccuracy itself will be unbounded.

Let $\beta_{\dot{m}}$ and $\beta_{\mathbf{v}}$ be positive scalars. Also, let $\|\cdot\|_2$ be the Euclidean norm, defined as $\|\mathbf{u}\|_2 = \left(\mathbf{u}^T\mathbf{u}\right)^{1/2}$ for a vector $\mathbf{u}$. One can state a context, C, for the model in Equation (4.1) as

$$C = \left\{\dot{m}(t), \mathbf{v}(t) : \left|\dot{m}(t)\right| \le \beta_{\dot{m}}, \left\|\mathbf{v}(t)\right\|_2 \le \beta_{\mathbf{v}}\right\}$$

where $\beta_{\dot{m}}$ and $\beta_{\mathbf{v}}$ bound the context for the time-derivative of mass and particle velocity, respectively. Thus, C is a set of world states defined by restrictions on the magnitudes of $\dot{m}(t)$ and $\mathbf{v}(t)$.

With a well-defined context, one can characterize inaccuracy. For this example, a set-based approach to representing inaccuracy and an additive inaccuracy model are adopted. This has the form:

$$\mathbf{F}(t) = m\mathbf{a}(t) + \mathbf{e}$$

where $\mathbf{e}$ is a vector inaccuracy term. The inaccuracy of the model is expressed as a bound on $\mathbf{e}$. One can develop more complex inaccuracy models. Ben-Haim describes several examples, including energy-bound, Minkowski-norm, slope-bound and Fourier-bound models (Ben-Haim 2001).

Assuming that the model in Equation (4.2) is perfectly accurate, one can compute the inaccuracy term as the difference between Equations (4.1) and (4.2). Thus, one has

$$\mathbf{e} = \dot{m}(t)\,\mathbf{v}(t),$$

which has a magnitude of

$$\varepsilon = \|\mathbf{e}\|_2 = \dot{m}\,\|\mathbf{v}\|_2 .$$

As discussed in Chapter 3, the ideal bound for a set-based inaccuracy representation is the least upper bound, or supremum. Here, the supremum is the maximum of the inaccuracy magnitude over the context set, or

$$\varepsilon_{\text{sup}} = \max_{C}\left(\|\mathbf{e}\|_2 = \dot{m}\,\|\mathbf{v}\|_2\right)$$
$$\varepsilon_{\text{sup}} = \beta_{\dot{m}}\beta_{v}$$

A validity description for the model in Equations (4.1) is summarized in Table 4.1. Although the model itself is not strictly a part of the validity description, it is repeated with the context and inaccuracy for easy reference. Essentially, one can think of the table contents as what model users would find in an appropriate repository. However, model users often will not have access to the model definition. Instead, they treat it as a "black box" with particular inputs and outputs. The statements in the context section of the table are interpreted as conditions that define a test for whether a situation is in the context. A situation must meet each of the stated conditions. The inaccuracy statement defines how one interprets the inaccuracy term and the limit of its magnitude for the associated context.

Note that the validity description of Table 4.1 is generalized for any choice of context bounds. Examples of validity descriptions for several specific contexts are given in Table 4.2. While in this case one can state the general validity description in closed form, this will not always be possible. In those cases, validity descriptions like those in Table 4.2 are reasonable.

4.2.2 Compatibility and Adequacy Assessment

Given a validity description, it is possible to assess the compatibility and adequacy of the associated model for an intended use based upon the characteristics of that intended use. As a demonstration of this, assume model users must solve the problem described in Figure 4.3. Furthermore, assume that a model is chosen with a validity description as given in Table 4.3. The first question to address is whether the model and problem are

Table 4.1: Summary of a generalized validity description for a formulation of Newton's second law of motion.

Model	$\mathbf{F}(t) - m\dot{\mathbf{v}}(t) = \mathbf{0}$	
Validity Description	Context	$\lvert \dot{m}(t) \rvert \le \beta_{\dot{m}}$ $\lVert \mathbf{v}(t) \rVert \le \beta_{\mathbf{v}}$
	Inaccuracy	$\mathbf{e} = \mathbf{F}(t) - m\dot{\mathbf{v}}(t)$ $\lVert \mathbf{e} \rVert_2 \le \beta_{\dot{m}}\beta_{\mathbf{v}}$

Table 4.2: Specific validity descriptions for a selection of contexts.

Model	$\mathbf{F}(t) - m\dot{\mathbf{v}}(t) = \mathbf{0}$			
Validity Description	Context	$\lvert \dot{m}(t) \rvert \le 10^{-9}$ $\lVert \mathbf{v}(t) \rVert \le 10^{3}$	$\lvert \dot{m}(t) \rvert \le 10^{-6}$ $\lVert \mathbf{v}(t) \rVert \le 10^{2}$	$\lvert \dot{m}(t) \rvert \le 10^{-1}$ $\lVert \mathbf{v}(t) \rVert \le 10^{4}$
	Inaccuracy	$\mathbf{e} = \mathbf{F}(t) - m\dot{\mathbf{v}}(t)$ $\lVert \mathbf{e} \rVert_2 \le 10^{-6}$	$\mathbf{e} = \mathbf{F}(t) - m\dot{\mathbf{v}}(t)$ $\lVert \mathbf{e} \rVert_2 \le 10^{-4}$	$\mathbf{e} = \mathbf{F}(t) - m\dot{\mathbf{v}}(t)$ $\lVert \mathbf{e} \rVert_2 \le 10^{3}$

An uncertain force, $\mathbf{F}_{\text{ext}}$, acts on a system. Assuming

$$\mathbf{F}_{\text{ext}}(t) = \begin{bmatrix} f_x(t) & 0 & 0 \end{bmatrix}^T \text{ N},$$

$$m_s = 500 \text{ kg},$$

$$|\dot{m}(t)| \leq 10^{-9} \text{ kg},$$

$$\|\mathbf{v}(t)\| \leq 30 \text{ m/s}$$

and the maximum allowable absolute inaccuracy is 10^{-8}, what is the acceleration of the system, $\dot{\mathbf{v}}(t)$?

Figure 4.3: Simulation Problem for compatibility and adequacy assessment example.

Table 4.3: Validity description for a specific context.

Model	$\mathbf{F}(t) - m\dot{\mathbf{v}}(t) = \mathbf{0}$	
Validity Description	Context	$\lvert\dot{m}(t)\rvert \leq 10^{-9}$ $\|\mathbf{v}(t)\| \leq 10^{2}$
	Inaccuracy	$\mathbf{e} = \mathbf{F}(t) - m\dot{\mathbf{v}}(t)$ $\|\mathbf{e}\|_2 \leq 10^{-7}$

context-compatible. Moreover, one first must perform compatibility assessment. Note that the model is included in the table only as a reference for readers. The assessment steps only make use of the validity description.

Compatibility Assessment

As stated in Figure 4.3, the desired behavioral attribute is the instantaneous acceleration, $\dot{\mathbf{v}}(t)$. One can read the context for this behavioral attribute (and the entire problem) directly from the problem statement. In this case, both the behavioral attribute context and the model context are specified in terms of bounds on the same variables. One can perform compatibility assessment by direct comparisons of the bounds. As discussed in Chapter 3, a model is context-compatible with a problem if its context subsumes that of the problem. For the adopted representations, this means that

$$(\|\mathbf{v}\|_2)_{\text{problem}} \leq (\|\mathbf{v}\|_2)_{\text{model}}$$

and

$$(|\dot{m}|)_{\text{problem}} \leq (|\dot{m}|)_{\text{model}}.$$

In this case, one finds that the model is context-compatible with the problem statement. The comparison is summarized in Table 4.4.

Table 4.4: Summary of results from compatibility assessment.

Quantity	Problem Context	Model Context	Compatible?
$\lvert \dot{m}(t) \rvert$	10^{-9}kg/sec	10^{-9} kg/sec	Yes
$\lVert \mathbf{v}(t) \rVert_2$	30 m/sec	100 m/sec	Yes
Final Assessment:	They are context-compatible.		

Adequacy Assessment

Given the model is context-compatible with the problem, one can perform adequacy assessment. The problem requires predictions of the acceleration to have an inaccuracy of no more than 10^{-6}. To determine whether the model is adequate, one can propagate model uncertainty through to the prediction assuming all other problem variables to be perfectly accurate. In this way, one can determine whether it is possible to make sufficiently accurate predictions using this model.

Reformulating the model to propagate the inaccuracy through to the acceleration, one obtains

$$\dot{\mathbf{v}}(t) = \frac{\mathbf{F}_{\text{ext}}(t) - \mathbf{e}}{m}$$

Because the force is directed along only one axis, one can reformulate this as a scalar problem. Doing so results in

$$\begin{aligned} \dot{v}_x(t) &= \frac{f_x(t)}{m} - \frac{e_x}{m} \\ &= \frac{f_x(t)}{m} - e_{\dot{v}_x} \end{aligned}$$

where e_x is the element of $\mathbf{e}$ corresponding to the x-dimension and $e_{\dot{v}_x}$ is an inaccuracy term for particle acceleration. Since $\max|e_x| \le \max(\|\mathbf{e}\|_2)$ one has

$$\max\left|e_{\dot{v}_x}\right| = \frac{\max|e_x|}{m} \le \frac{\max(\|\mathbf{e}\|_2)}{m} = \frac{10^{-7}}{500} = 2\times 10^{-9}.$$

This is the prediction inaccuracy due to model inaccuracy. Since this inaccuracy is smaller than the maximum allowable prediction inaccuracy of 10^{-8}, one can declare the model to be adequate. Moreover, the model is valid for this particular use.

4.2.3 Remarks

Drawing upon the preceding example, one can make several observations about validity characterization and compatibility and adequacy assessment. They are as follows.

Context Definition

In this example, model context is defined by terms not actually in the model, namely $\dot{m}$ and $\mathbf{v}$. This happens when an analyst makes simplifications during modeling. When terms are eliminated from a model because they are assumed "insignificant," they must be bounded in its context. Specifying bounds on these terms in a context defines the semantics of the assumption. That is, it defines what it means to be "insignificant".

Context-Inaccuracy Relationship

One can observe a relationship between context and inaccuracy in this example. Expanding the context (i.e., raising the bound on either or both context variables) results in an increased inaccuracy. This is because the inaccuracy is defined in terms of the context variables.

In general, an expansion of a context cannot result in a decrease in inaccuracy, and often will result in an increase. For the single-parameter case illustrated above, this means that

$$C_1 \subset C_2 \rightarrow e_1 \leq e_2$$

where C_1 and C_2 are contexts and ε_1 and ε_2 are the corresponding inaccuracy parameters.

<u>Correctness of Inaccuracy Bound</u>

Strictly speaking, the validity characterization performed in this model is incomplete. It fails to capture all of the inaccuracy in the model because Equation (4.2) is not perfectly accurate. Thus, the inaccuracy stated in the validity description reflects the relative inaccuracy between the models in Equations (4.1) and (4.2). The best generally accepted model for the force-acceleration relationship is Einstein's theory of special relativity, which implies that that (Ohanian 1995)

$$\mathbf{F}(t) = \frac{d}{dt}\left(\frac{m(t)\,\mathbf{v}(t)}{\sqrt{1 - {v^2(t)}/{c^2}}}\right)$$

where $v(t) = \left\|\mathbf{v}(t)\right\|_2$ is the particle speed and c is the speed of light in a vacuum. This model is nearly identical to that of Equation (4.2) for particle speeds that are not a significant fraction of the speed of light. This is the case in the example presented in this section. However, the inaccuracy can grow to be significant in other contexts.

In principle, the objective of validity characterization is to identify all of the known inaccuracy within a given context. To accomplish this, one must turn to the best known empirical methods or theoretical models (i.e., a model that is regarded as "ground truth" or "first principles" by experts in that domain). In practice, performing validity characterization relative to first principles may require exceptional effort while yielding only modest expansion of the inaccuracy. Model creators must use their understanding of a problem domain to determine an appropriate and useful validity description. It is their responsibility to specify conservative bounds on inaccuracy in order to prevent misuse of the models they create.

Empirical Performance Evaluation of Hypotheses

The hypotheses are useful for this example problem. Given the assumed methods and representations, model creators and model users can perform model validation of a reusable behavioral model by performing the steps specified in the hypotheses. The flow of steps in the example corresponds to process flow depicted in Figure 3.4. This proves to be effective.

A main feature of the proposed hypotheses is that they specify what knowledge model creators and model users must communicate so that model users can validate the use of a model. This knowledge comprises a validity description for a model. The only knowledge communicated between the model development and model use processes in this example is contained in validity description of Table 4.3. Thus, the validity description is successful as a conduit of validation-relevant knowledge.

4.3 An Example using a Model of Beam Extension under Axial Tension

This example consists of the development and validity characterization of a model for a structural steel beam held in axial tension. It is a demonstration of validity characterization as a part of model development. The model of interest is more complex than that of the previous example, but is not so complex that validity characterization is overshadowed by model development.

Model development and validity characterization proceeds in two steps. First, a validity characterization is developed for the stress-strain relationship for a homogenous material. This relationship typically is referred to as Hooke's law. Given the results of this step, a model is developed for static strain in an axially loaded beam. A validity description is developed for this model based upon the one for Hooke's law and knowledge of the physical scenario being modeled.

Section 4.3.1 contains background material about strain and the factors that can influence it. Section 4.3.2 contains an account of the characterization of Hooke's law based upon knowledge of thermal strain effects. Section 4.3.3 contains a description of the target model scenario and an account of the development and validity characterization of an appropriate model. Section 4.3.4 contains remarks about this example.

4.3.1 Preliminaries

<u>Stress and Strain</u>

In mechanical engineering, the notion of stress relates an applied force to an area normal to the force. There are two common definitions of stress. *Engineering stress* is defined as

$$\sigma = \frac{F}{A_0},$$

where F is the applied force and A_0 is the cross-sectional area prior to application of the force (i.e., the initial area). *True stress* is defined as

$$\tilde{\sigma} = \frac{F}{A},$$

where F is the applied force and A is the cross-sectional area after application of the force (i.e., the final area). Both definitions lead to similar stress values when deformations are small. The true stress is more accurate for larger changes in area.

Strain is defined as the elongation per unit length. As with stress, there are two common definitions for strain. *Engineering strain* is defined as

$$\varepsilon = \frac{\Delta L}{L_0},$$

where L_0 is the initial material length and ΔL is the change in material length. *True strain* is based on a differential definition. The differential strain for a change in length, dl, over a length, l, is defined as

$$d\tilde{\varepsilon} = \frac{dl}{l}.$$

One obtains the total true strain through integration. For a material of initial length L_0 and final length L, one has

$$\tilde{\varepsilon} = \int_{L_0}^{L} \frac{dl}{l} = \ln\left(\frac{L}{L_0}\right).$$

Similar to the stress definitions, the true strain is more accurate at larger deformations.

Hooke's Law

For many materials, a plot of stress versus strain results in a straight line for a significant range of stress values. The linear stress-strain relationship in this region is referred to as Hooke's law and is stated as:

$$\sigma = E\varepsilon \tag{4.3}$$

where E, the slope of the line, is referred to as the Young's modulus of the material. This relationship is independent of whether one uses engineering or true stress and strain definitions. It holds for bars loaded in tension along a single axis and assumes that other factors that influence strain—such as temperature—are kept constant. Although not considered here, one can generalize this relationship to include stresses and strains along all axes.

Thermal Strain

Most materials experience strains due to changes in temperature. It is common in design to approximate thermal strain as a linear function of temperature change. Thus, we have

$$\varepsilon_T = \alpha\left(T_f - T_i\right),$$

where T_f is the final temperature, T_i is the initial temperature and α is known as the coefficient of thermal expansion and depends on the material. Although more complex and precise relationships for this effect exist, this relationship is assumed perfectly accurate in this example.

In general, this relationship is accurate for a small range of possible temperatures. Outside of this context, the relationship becomes inaccurate and eventually breaks down altogether. For instance, the material in question will melt if the temperature becomes too high. Let T_{LB} and T_{UB} be, respectively, lower and upper bounds on allowable temperature values for the material of interest. In this context, the magnitude of the maximum temperature difference is $\Delta T_{\max} = T_{UB} - T_{LB}$.

Total Strain

The total strain is the sum of the thermal strain and the elastic strain:

$$\varepsilon_{\text{total}} = \varepsilon_{\sigma} + \varepsilon_{T} = \frac{\sigma}{E} + \alpha\left(T_f - T_i\right). \tag{4.4}$$

4.3.2 Validity Characterization of Hooke's Law

Presented in this subsection is the validity characterization of Hooke's law (Equation (4.3)). The present analysis is limited to strain along one axis under the assumption that this leads to no additional inaccuracy. As presented here, the inaccuracy of Equation (4.3) is due to thermal effects. Structural steel is the material of interest. Selection of a different material would change the outcome, but the process would remain the same in principle.

Inaccuracy Representation Approaches

In characterizing this model, data for a particular material are assumed available. Figure 4.4 contains a notional depiction of what this data might look like (note: scale of noise in data is exaggerated relative to scale of graph). While designers treat the relationship as holding perfectly, the data supports only that the relationship holds to within some degree of accuracy. This inaccuracy can be represented in multiple ways. Two possibilities are as follows:

First, one can treat Young's modulus, E, as an uncertain quantity that has an upper and lower bound (or a probability distribution if there is sufficient evidence to support this). This approach assumes that the relationship itself is correct for the given data. Thus, while the relationship itself contributes no inaccuracy, any computation with the relationship would also include the uncertain parameter, E, that would introduce inaccuracy.

The second approach is to select the "best fit" slope for the relationship and assume this as the value for E. Then, the relationship is considered to have inaccuracy, but we presume to know the Young's modulus exactly.

The impact of the two different approaches is depicted in Figure 4.5. By embodying the inaccuracy in the slope parameter (Young's modulus), the overall inaccuracy is lower at small stress values. However, a similar effect can be achieved by using an inaccuracy representation that varies with stress. For the current example, the approach shown in Figure 4.5(a) is adopted. Thus, noise in the data contributes no inaccuracy to the model, but manifests itself in a prediction by being propagated from a model parameter.

Accounting for Thermal Strain Effects

A change in temperature results in a strain in a material. By neglecting this effect, Hooke's law is inaccurate relative to the true stress-strain relationship. To gauge the inaccuracy, one can determine the error introduced by neglecting thermal effects.

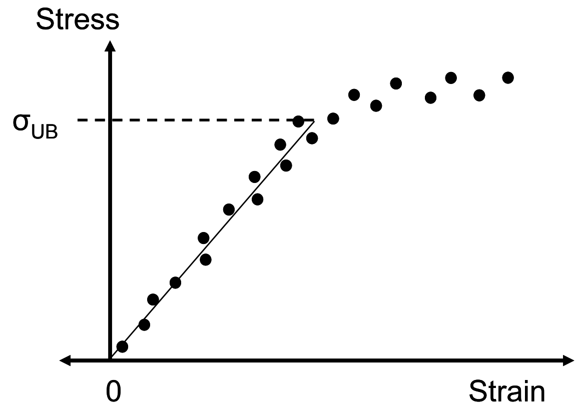

Figure 4.4: A notional depiction of stress-strain sample points for a particular material. The "elastic region" is the nearly-linear part of the data located below σ_{UB}. Note that noise in the data is exaggerated relative to the scale of the graph.

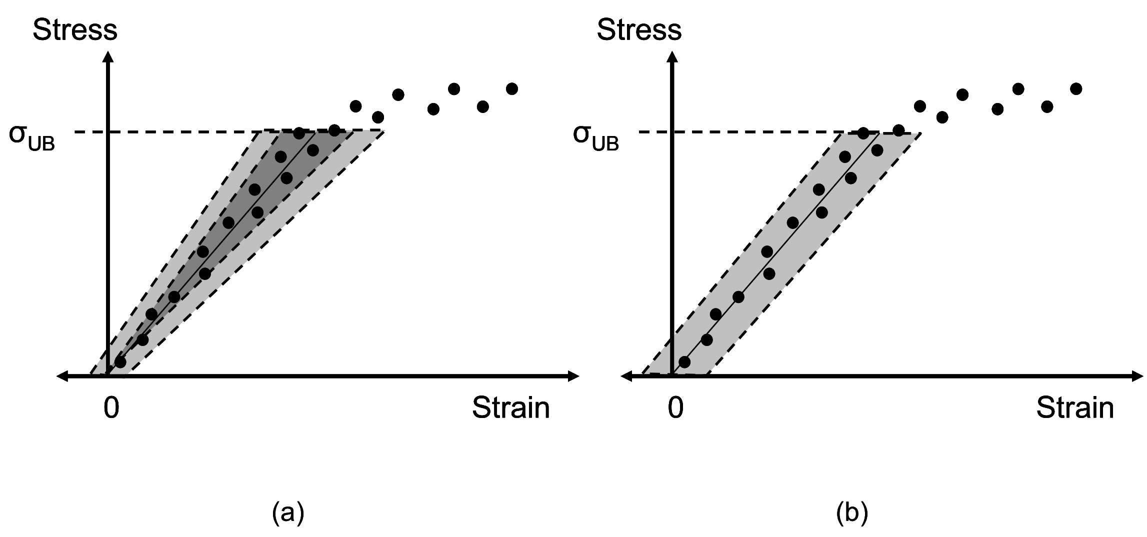

Figure 4.5: A conceptual comparison of the implications of representing inaccuracy: (a) in the quantity E , and (b) in the Hooke's law relationship. In (a), the dark grey region represents inaccuracy in the parameter E. In both, the light grey region represents model inaccuracy.

Substituting the total strain from Equation (4.4) into Hooke's law (Equation (4.3), one has:

$$\sigma = E\left(\varepsilon_{\sigma} + \varepsilon_{T}\right)$$
$$\sigma - E\varepsilon_{\sigma} = \alpha\left(T_{f} - T_{i}\right)$$

or

$$\delta_{Hooke} = \alpha E\left(T_{f} - T_{i}\right) \tag{4.5}$$

where δ_{Hooke} is the error in Hooke's law due to thermal strain. It is a function of material properties (thermal expansion coefficient and Young's modulus) and environmental conditions (temperature). A bound on this term, β_{Hooke}, serves as an inaccuracy representation for this example. The next step is to determine a context over which to evaluate this bound.

Context

For this example, the context (and thus the corresponding validity description) is for a particular material: structural steel. Thus, the selection of context restrictions depends on the known material properties for that material (which presumably result from empirical test data).

Context formalization proceeds in two steps: identification of terms that require bounding and selection of specific bound values.

In order to establish a bound, β_{Hooke}, for δ_{Hooke} (i.e., to state the inaccuracy), one must define a context that involves the terms in Equation (4.5) along with any others that influence model inaccuracy. Bounds on the Young's modulus are based on the material properties of structural steel. Let E_{UB} and E_{LB} be the upper and lower bounds, respectively. Thus, $E \in \left[E_{LB}, E_{UB}\right]$. The coefficient of thermal expansion is positive and assumed bounded on some range, defined as $\alpha_{LB} \leq \alpha \leq \alpha_{UB}$. These bounds also result from the material properties of structural steel.

The initial and final temperatures require bounds, but these are not particular to the material. For these, model creators must use their judgment to select values that are appropriate for model users. Bounds for the temperature variables are discussed in Section 4.3.1. The result is that

$$\Delta T_{\max} = \max\left|T_f - T_i\right| = T_{UB} - T_{LB}$$

where T_{LB} and T_{UB} are, respectively, lower and upper bounds on T_f and T_i.

One also must assume bounds on stress or strain. This bound is necessary because the nearly-linear stress-strain relationship breaks down at large stresses and strains. This is depicted in Figure 4.4. For this example, stress is assumed to be positive (i.e., the beam is never in compression) and less than an upper bound, σ_{UB}. This yields a corresponding restriction on strain, with a lower bound of zero and upper bound of ε_{UB}. Thus, $\sigma \in \left[0, \sigma_{UB}\right]$ and $\varepsilon \in \left[0, \varepsilon_{UB}\right]$. Since stress and strain are functionally related through the model, a bound on one value establishes a bound on the other (assuming bounds exist on the Young's modulus). One can define bounds on both. In such situations, the more restrictive of the two bounds dominates. For instance, one might define a stress bound such that corresponds to strains that are less than the strain bound (i.e., ${}^{\sigma_{UB}}\!/_{E_{LB}} < \varepsilon_{UB}$). There is no harm is specifying an unreachable bound. The main concern is specifying a context for which the inaccuracy is reasonable. Whenever possible, one should specify a context in terms of model inputs. This allows model users to perform compatibility assessment without first simulating the model.

The selection of numerical values for the bounds is at the discretion of model creators. As noted in Chapter 3, inaccuracy tends to increase as a context expands. This suggests that model creators may wish to select restrictive context bounds. However, a narrow context limits the opportunities for reuse. This leaves model creators with a tradeoff they must negotiate as best they can using their domain knowledge and any available information about probable future model uses.

The numerical bounds for this example are stated in Table 4.5. They are representative of structural steel, but may apply to other materials. The upper bound on stress is the yield strength of ASTM-A36 structural steel (Halliday, et al. 1988). This is the point at which the material becomes inelastic. The error in Hooke's law relative to

Table 4.5: Numerical values for Hooke's law context bounds.

Variable	Symbol	Lower Bound	Upper Bound	Units
Coefficient of Thermal Strain	α	$\alpha_{LB} = 11\times10^{-6}$	$\alpha_{UB} = 13\times10^{-6}$	$1/\mathrm{K}$
Stress	σ	0	$\sigma_{UB} = 250$	MPa
Strain	ε	0	$\varepsilon_{UB} = 0.011$	
Young's Modulus	E	$E_{LB} = 180$	$E_{UB} = 220$	GPa
Temperature	T_f, T_i	$T_{LB} = 290$	$T_{UB} = 300$	K

empirical data grows significantly beyond this point. The range for Young's modulus is defined as $\pm 10\%$ about a typical value for structural steel. The range for the coefficient of thermal strain is based on a similar range about a typical value for structural steel. The temperature range covers small variation about room temperature.

Inaccuracy Bound

One can approximate a bound on the inaccuracy of the Hooke's law model by considering the impact of the assumptions embodied in the model. The model formulation of Equation (4.3) includes the assumption that thermal expansion is insignificant. One can evaluate the impact this assumption has on model accuracy using the relationship of Equation (4.5) and the context bounds from Table 4.5:

$$\begin{aligned}
\delta_{\text{Hooke}} &\le \beta_{\text{Hooke}} = \max\left|\alpha E\left(T_f - T_i\right)\right| \\
&\le \beta_{\text{Hooke}} = \alpha_{UB} E_{UB} \Delta T_{\max} \\
&\le \beta_{\text{Hooke}} = \left(13\times 10^{-6}\right)\left(220\times 10^{9}\right)\left(300-290\right) \\
&\le \beta_{\text{Hooke}} = 28.6 \text{ MPa}
\end{aligned}$$

Thus, the inaccuracy of Hooke's law due to neglecting thermal expansion is no more than 28.6 MPa anywhere within the stated context. This is about 11% of the maximum stress within the context. It serves as an approximate least upper bound on model inaccuracy.

Validity Description

To summarize, validity characterization of the Hooke's Law model results in the validity description given in Table 4.6. It is a formal description of a model creator's

Table 4.6: A validity description for Hooke's law.

Model	$\sigma - E\varepsilon = 0$	
Validity Description	Context	$\sigma \in [0,250]$ MPa, $\varepsilon \in [0,0.011]$, $T_f, T_i \in [290,300]$ K, $\alpha \in [11,13]$ $10^{-6}/\mathrm{K}$, $E \in [180,220]$ GPa
	Inaccuracy	$\delta_{\text{Hooke}} = \sigma - E\varepsilon$ $\lvert\delta_{\text{Hooke}}\rvert \leq \beta_{\text{Hooke}} = 28.6$ MPa

knowledge about the assumptions embodied in the model. Other validity descriptions for this model are possible. Model creators can arrive at them by defining a different context set or by using a different inaccuracy representation.

4.3.3 Validity Characterization of a Beam Under Tension

Target Scenario

The physical scenario is depicted (with the deformation greatly exaggerated) in Figure 4.6. The beam has initial length, L_0, and cross-sectional area, A_0. The beam in held in static tension by a load force, F, which is applied uniformly over the cross-sectional area. The beam is assumed to be massless and to have a constant cross-sectional area along its length. Structural steel is assumed as a material for the beam.

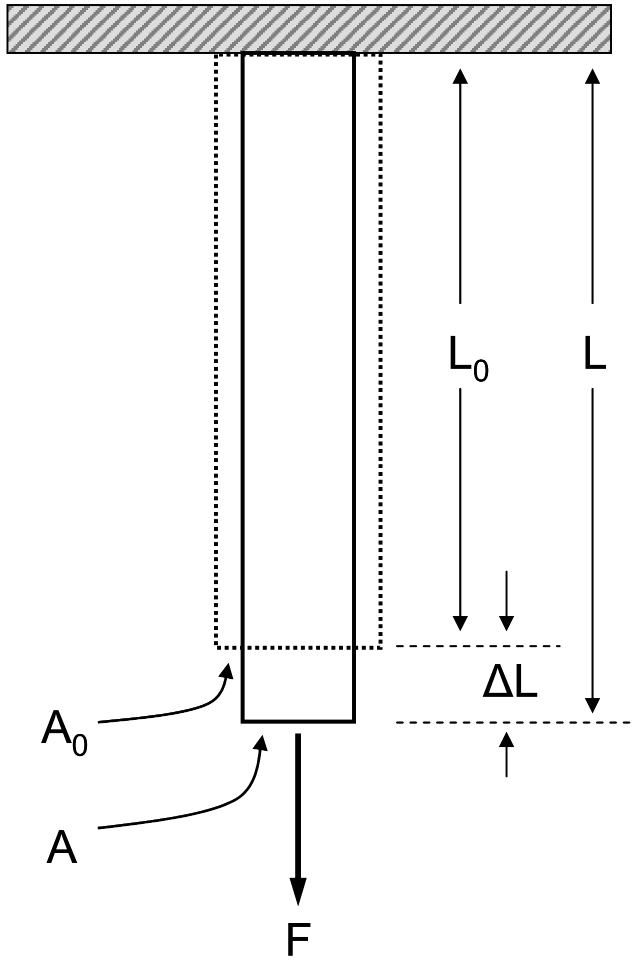

Figure 4.6: Depiction of beam in axial tension (deformation exaggerated).

Under these assumptions, the stress is uniform throughout the beam. Substituting the definitions for engineering stress and engineering strain into the Hooke's law model from the previous subsection (Equation (4.3)), yields:

$$\begin{aligned} \sigma &= E\varepsilon \\ \frac{F}{A_0} &= E\frac{\Delta L}{L_0} \end{aligned}$$

Some rearranging yields an expression for the change in length:

$$\Delta L = L_0 \frac{F}{A_0 E}. \tag{4.6}$$

This is the model for which validity characterization is performed in this subsection.

Models with Weaker Assumptions

In order to characterize the model (Equation (4.6)), one must establish a baseline for comparison—i.e., a referent. For the present example, a model with less restrictive assumptions serves as a referent. This is similar to the approach taken for the Hooke's law model from Section 4.3.2 and the Newton's law model from Section 4.2. Other referents might include empirical data or the knowledge of domain experts.

The model of Equation (4.6) includes many assumptions. The impact of two particular assumptions are examined in detail here. The assumptions are that:

- The difference between the engineering and true definitions of stress and strain is negligible.
- Thermal expansion is negligible.

It is possible to gauge the impact of these assumptions based on material from earlier in this section.

Table 4.7: Engineering and true definitions for stress and strain.

	Stress	Strain
Engineering	$\sigma = \dfrac{F}{A_0}$	$\varepsilon = \dfrac{\Delta L}{L_0}$
True	$\tilde{\sigma} = \dfrac{F}{A}$	$\tilde{\varepsilon} = \ln\left(\dfrac{L}{L_0}\right)$

Section 4.3.1 includes two definitions for each stress and strain. One pair of definitions is referred to as the "engineering" definitions. They are good approximations at small deformations and are convenient because most of the information required to compute with them is based on the initial conditions of the system. The "true" definitions are more accurate over a wider range of circumstances. Table 4.7 is a summary of the different definitions, where F is a force applied to a beam of initial cross-sectional area A_0 and initial length L_0. The final cross-sectional area is A and final length is L. The change in length is defined as $\Delta L = L - L_0$.

Using the definition for true strain, one can derive an expression for the change of beam length as a function of strain. One has

$$\begin{aligned} \tilde{\varepsilon} &= \ln\left(\frac{L}{L_0}\right) \\ e^{\tilde{\varepsilon}} &= \frac{L}{L_0} \\ L_0 e^{\tilde{\varepsilon}} &= L \end{aligned}$$

that, when combined with the definition for ΔL, yields:

$$\Delta L = L - L_0$$
$$\Delta L = L_0 e^{\tilde{\varepsilon}} - L_0$$

or

$$\Delta L = L_0 \left(e^{\tilde{\varepsilon}} - 1 \right). \tag{4.7}$$

To serve as a referent for characterizing Equation (4.6), one must reformulate this expression in terms of force instead of strain. One can accomplish this by applying Hooke's law and the definition for true stress. However, Hooke's law contains inaccuracy. One must account for this inaccuracy when developing the beam model referent. To neglect it would result in an improper validity description that does not include all of a model creator's knowledge about the model.

According to the validity description from Table 4.6,

$$\tilde{\sigma} - E\tilde{\varepsilon} = \delta_{\text{Hooke}},$$

where $\delta_{\text{Hooke}} \in \left[-\beta_{\text{Hooke}}, \beta_{\text{Hooke}} \right]$. One can rearrange this and substitute the definition for true stress (see Table 4.7) to achieve:

$$\tilde{\varepsilon} = \frac{1}{E}\left(\frac{F}{A} - \delta_{\text{Hooke}} \right)$$

where A is the final cross-sectional area of the beam. Substituting this into Equation (4.7) yields:

$$\Delta L = L_0 \left(e^{\frac{1}{E}\left(\frac{F}{A} - \delta_{\text{Hooke}} \right)} - 1 \right). \tag{4.8}$$

This model serves as the referent for the validity characterization of Equation (4.6).

Context

Because it is derived using Hooke's law, the context for the beam model of Equation (4.6) inherits the Hooke's law context restrictions. However, one must include additional restrictions because the beam model involves assumptions beyond those of the Hooke's law model. One must define a context such that the bound on inaccuracy is finite within the context. For this problem, inaccuracy is determined using the relationships in Equations (4.6) and (4.8). Thus, one must define context restrictions on all the variable appearing in that expression. In addition to the variables restricted by the Hooke's law context, the force, beam length and cross-sectional area require context restrictions.

Table 4.8 is a summary of the chosen context bounds along with those inherited from the Hooke's law context. The beam dimensions are representative of a long, slender rod. The bounds for the applied force result in stresses that are well within the context restrictions for stress.

Inaccuracy

One can approximate the inaccuracy for the beam model of Equation (4.6) by considering the impact of its assumptions. Equation (4.8) can serve as a referent for the purpose of appraising the inaccuracy of the beam model. For this example, a bound on the absolute difference between the model and the referent represents the inaccuracy. Let

$$\Delta L - L_0 \frac{F}{A_0 E} = \delta_{\mathrm{beam}}$$

where δ_{beam} is the error in the model. The inaccuracy bound, β_{beam}, is

Table 4.8: Summary of context restrictions for model of beam in axial tension.

Variable	Symbol	Lower Bound	Upper Bound	Units
Coefficient of Thermal Strain	α	$\alpha_{LB} = 11\times10^{-6}$	$\alpha_{UB} = 13\times10^{-6}$	$1/\mathrm{K}$
Stress	σ	0	$\sigma_{UB} = 250$	MPa
Strain	ε	0	$\varepsilon_{UB} = 0.011$	
Young's Modulus	E	$E_{LB} = 180$	$E_{UB} = 220$	GPa
Temperature	T_f, T_i	$T_{LB} = 290$	$T_{UB} = 300$	K
Length	L_0, L	$L_{LB} = 0.4$	$L_{UB} = 0.6$	m
Cross-sectional Area	A_0, A	$A_{LB} = 130$	$A_{UB} = 150$	mm^2
Applied Force	F	$F_{LB} = 10$	$F_{UB} = 100$	kN

$$
\begin{aligned}
\delta_{\text{beam}} &= L_0\left(e^{\frac{1}{E}\left(\frac{F}{A}-\delta_{\text{Hooke}}\right)}-1\right)-L_0\,\frac{F}{A_0E} \\
&\le \beta_{\text{beam}} = \max\left|L_0\left(e^{\frac{1}{E}\left(\frac{F}{A}-\delta_{\text{Hooke}}\right)}-1\right)-L_0\,\frac{F}{A_0E}\right| \\
&\le \beta_{\text{beam}} = \max\left|L_0\left(e^{\frac{1}{E}\left(\frac{F}{A}-\delta_{\text{Hooke}}\right)}-\frac{F}{A_0E}-1\right)\right| \\
&\le \beta_{\text{beam}} = L_{UB}\left(e^{\frac{1}{E_{LB}}\left(\frac{F_{UB}}{A_{LB}}+\beta_{\text{Hooke}}\right)}-\frac{F_{UB}}{A_{LB}E_{LB}}-1\right) \\
&\le \beta_{\text{beam}} = 0.24\ mm
\end{aligned}
$$

Thus, the inaccuracy in the beam extension model is about a quarter of a millimeter over the context specified in Table 4.8. This represents less than 10% of the maximum ΔL

prediction possible within the stated context. This is an approximate bound on inaccuracy. One might be able to find a tighter bound relative to this referent. However, other inaccuracies are not accounted for in the referent. For instance, the referent maintains the assumption that the mass of the beam is insignificant.

Validity Description

The corresponding validity description is given in Table 4.9. It is a formal representation of a model creator's validation-relevant knowledge. This validity description is fairly restrictive. The context is less wide than that for Hooke's law. This is at the discretion of the model creator. One can develop a different validity description for this model as needs and desires dictate.

4.3.4 Remarks

Demonstrated in this example is the validity characterization of a model for an elastic beam held statically in axial tension. The model is developed and characterized in two steps. In the first step, a validity description is developed for the stress-strain relationship in a homogenous elastic material, commonly referred to as Hooke's law. In the second step, the Hooke's law model and its validity description are used as building blocks for the development and validity characterization of the beam model. The following remarks are in order.

Table 4.9: A validity description for the axially loaded beam example problem.

Model		$\Delta L = L_0 \dfrac{F}{A_0 E}$
Validity Description	Context	$\sigma \in [0, 250]$ MPa, $\varepsilon \in [0, 0.011]$, $T_f, T_i \in [290, 300]$ K, $\alpha \in [11, 13]\ ^{10^{-6}}\!/_{\mathrm{K}}$, $E \in [180, 220]$ GPa $L_0, L \in [0.4, 0.6]$ m $A_0, A \in [130, 150]$ mm^2 $F \in [10, 100]$ kN
	Inaccuracy	$\Delta L - L_0 \dfrac{F}{A_0 E} = \delta_{\mathrm{beam}}$ $\lvert \delta_{\mathrm{beam}} \rvert \le \beta_{\mathrm{beam}} = 0.24\ mm$

Relationships among Context Variables

In the first example, the context is defined solely in terms of variables that are not in the model. For the validity characterization of Hooke's law, both stress and strain appear in the model and in the context. This is necessary because the inaccuracy of the linear stress-strain model is a function of the location on the curve. This is depicted in Figure 4.4. The relationship becomes nonlinear beyond particular stress-strain values. Both stress and strain appear in the validity description because the model is non-causal. That is, a model user may select stress or strain to be the dependent variable (i.e., the prediction). In this example, the context bounds on stress and strain equate to the same points on the stress-strain curve. This is not strictly necessary. The context is defined as the intersection of the individual context constraints. As soon as one variable bound is violated, contexts are no longer compatible. For related context variables, the tightest variable bound trumps the others.

Inaccuracy Formulation

Although this example is relatively simple, it involves several important decisions. One decision is how to formulate the inaccuracy in Hooke's law. Inaccuracy formulations of two different set-based inaccuracy models is depicted conceptually in Figure 4.5. The decision is arbitrary for this example because the problem is synthetic. In practice one should choose the inaccuracy representation that is best supported by the available data or theory. In general, this may involve the use of a formalism other than set-theory, such as probability theory or any of the others discussed in Chapter 3.

Empirical Performance Evaluation

This example is a demonstration of validity characterization for a more complex model than the one used in Section 4.2. Although more complex, it is possible to construct a validity description that captures the requisite validation-relevant knowledge about the model. This is evidenced by the derivation of an inaccuracy bound over the stated context. This knowledge is shown to be sufficient for model users to perform validation in the example from the previous section. Thus, the central concept in the proposed framework—the validity description—proves useful on this example.

The reason for focusing on validity characterization in this example is that the output of this activity forms a basis for the subsequent assessment activities. If the resulting validity description does not contain the necessary validation relevant knowledge, one cannot proceed with the assessment steps. Given a formalized validity description and problem definition, one can perform compatibility and adequacy assessment relatively easily. They essentially are mechanical processes involving the evaluation of mathematical comparisons. In contrast, validity characterization is a creative process involving the expertise of model creators. As evidenced in this example, model creators must make decisions such as how to formulate inaccuracy and how large to make the context. These are expertise-driven decisions and they fit well within the proposed framework. What is more, one of the strengths of the framework is that it separates activities that require significant creativity and insight from those that tend to be more mechanical.

4.4 Summary

This chapter contains two examples of using the proposed conceptual framework. The example of Section 4.2 is a demonstration of the three primary model validation activities—validity characterization, compatibility assessment and adequacy assessment—on a basic model. The example of Section 4.3 involves validity characterization of a more complex model. Because the framework is abstract, particular methods and representations are adopted. These are summarized in Section 0.

The conceptual framework defined in the hypotheses proves useful on the example problems. The first example is conducted under the assumptions of a general model reuse scenario. Validity characterization is produced with no knowledge of future model uses. Compatibility and adequacy assessment are performed with no knowledge of the model other than what is contained in its validity description. Although the model is stated in Table 4.3, only the validity description is used during compatibility and adequacy assessment. This reflects the reality of most reuse scenarios, where model users interact with a model as a black box.

The focus of the second example is validity characterization. This example serves as an illustration of the concepts of context and inaccuracy and as a demonstration that one can define validity descriptions for more complex models. However, this model is still simple relative to many engineering models. A discussion of how the framework might extend to more complex problems is included in the next chapter. It also includes an overall evaluation of the hypotheses and discussion of potential areas for future work.

CHAPTER 5:

DISCUSSION AND REMARKS

This chapter is a discussion and synthesis of material from the preceding chapters. It involves an examination of the thesis from two perspectives. The first involves an inward reflection upon material from the previous chapters. Section 5.1 contains this discussion. The hypotheses are reviewed and evaluated according to the strategy outlined in Section 1.4.2. The second perspective is an outward perspective. This involves a consideration of the broader implications and limitations of the work, both in actuality and in its potential. Section 5.2 is a discussion of the contributions of the work. Section 5.3 is a discussion of the limitations of this work, both fundamental and as a result of thesis methodology. Section 5.4 is a discussion of how this work forms a basis for future investigations of behavioral modeling and simulation in engineering design. It includes descriptions of potential extensions of the work. Section 5.5 contains some closing remarks. Figure 5.1 includes a summary of the objectives of this chapter.

5.1 Review and Evaluation of Hypotheses

This section is a review and evaluation of the hypotheses proposed in this thesis. It marks the point of closure for the evaluation strategy outlined in Section 1.4.2. Section 5.1.1 contains a review of the research questions and proposed hypotheses. It is a summary of the detailed discussions of the preceding chapters. The remaining subsections are an account of the hypothesis evaluation efforts. Section 5.1.2 is a review

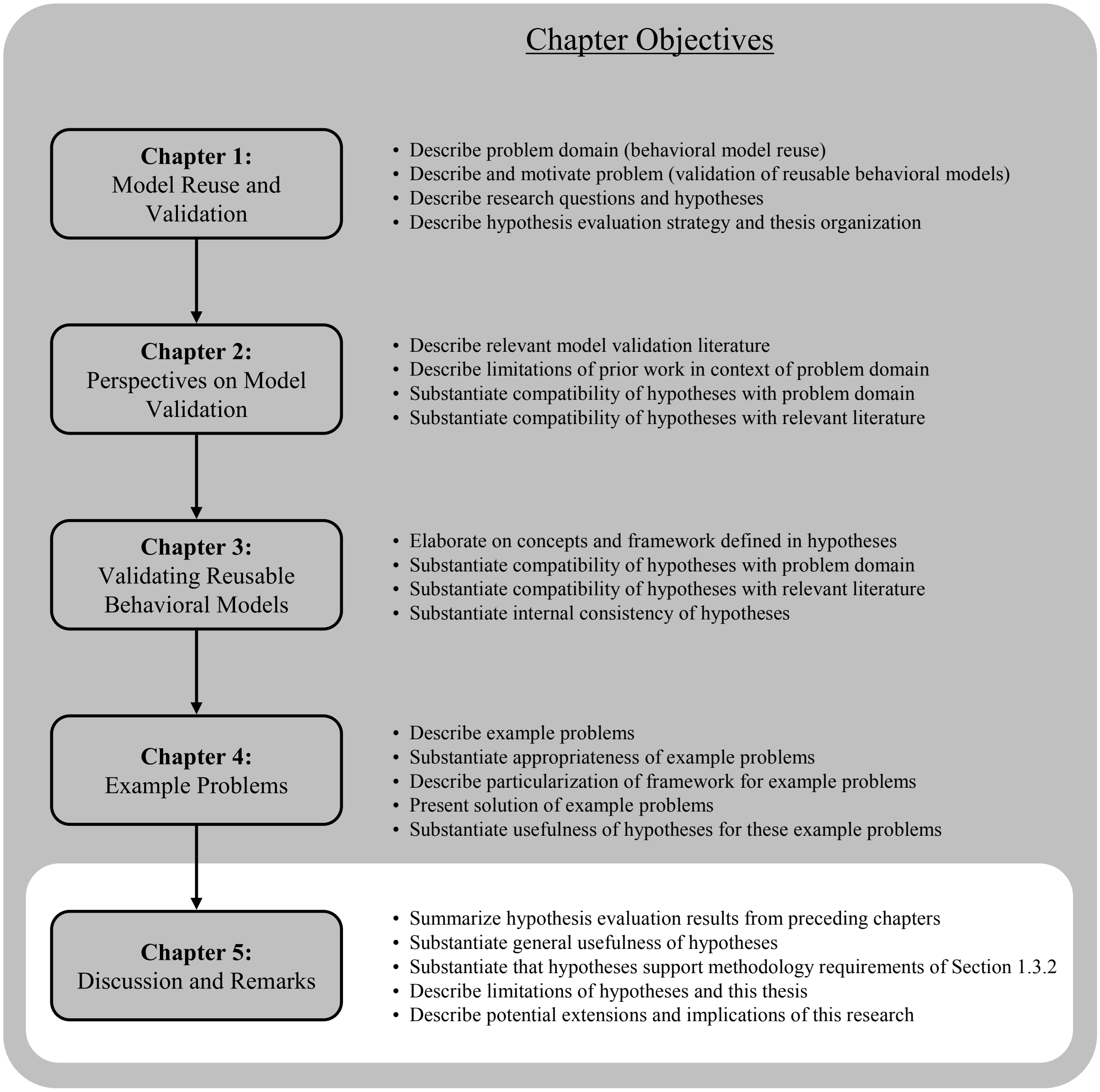

Figure 5.1: Summary of Chapter 5 objectives in relationship to rest of thesis.

and discussion of theoretical structural evaluation efforts. Section 5.1.3 is about empirical structural evaluation. Section 5.1.4 is about empirical performance evaluation. Section 5.1.5 covers theoretical performance evaluation.

5.1.1 Review of Research Questions and Hypotheses

Chapter 1 contains a discussion about behavioral model reuse in engineering design and the challenges of performing model validation in this setting. The main results are as follows:

- In order to perform validation on a behavioral model, one requires knowledge about the system being modeled, the limitations of the model and the objectives of the simulation study (Section 1.2.2).
- In typical behavioral model reuse scenarios, engineering designers who seek to reuse a model lack knowledge about the limitations of a model while the creators of the model lack knowledge about the objectives of the simulation study (Section 1.3).
- In several potential behavioral model reuse scenarios, model users and model creators may be unable to exchange knowledge by interacting directly (Section 1.3.1).

Thus, model reuse potentially leads to a knowledge deficiency that can impede or prevent model validation. This serves as the motivation for the primary research question of this thesis:

Primary Research Question: *How can engineering designers perform behavioral model validation in a way that supports model reuse?*

The primary research question is decomposed based on the observation that the potential knowledge deficiency is due to limitations on how model creators and model users interact. A result from the discussion of Section 1.3.2 is that model creators and model users must have a means to communicate validation-relevant knowledge in an efficient and unambiguous fashion. The flow of this knowledge in typical reuse scenarios is *from* model creators *to* model users. One can observe this in the reuse scenarios described in Section 1.3.1. Thus, the primary research question is decomposed according to how model creators can convey their knowledge to model users and how model users can apply this knowledge.

> ***Q1:*** *How can model creators convey validation-relevant knowledge in a way that is independent of any person, group or project?*
>
> ***H1:*** *Model creators can develop mathematical descriptions of their models—called validity descriptions—that provide assertions about the accuracy a user can expect and the context over which the assertions hold true*

Q2: *How can model users apply validity descriptions to validate the application of a behavioral models to a particular problem?*

H2: *Model users can perform a two-step assessment process in which they:*

(1) determine whether the context stated in the validity description is compatible with the problem and, if it does,
(2) determine whether the accuracy stated in the validity description is sufficient for the needs of the problem.

Together, these hypotheses comprise a conceptual framework for the validation of behavioral models that is compatible with model reuse scenarios in engineering design. They are elaborated throughout the thesis. Chapter 2 contains a discussion of how the hypotheses relate to prior work on model validation. Chapter 3 is an elaboration of validity descriptions, with a focus on the fundamental concepts that comprise them. Chapter 4 contains example problems that illustrate the conceptual framework and how one can extend it to solve validation problems.

Section 1.4.2 contains an outline of the hypothesis evaluation strategy adopted for this thesis. The strategy is summarized in Table 1.4 which is repeated below in Table 5.1 for convenience. The strategy involves four steps: theoretical structural evaluation, empirical structural evaluation, empirical performance evaluation and theoretical performance evaluation. With the exception of theoretical performance evaluation, results and discussion relevant to hypothesis evaluation appear throughout the preceding

Table 5.1: Summary of hypothesis evaluation strategy from Table 1.4.

Evaluation Step	Description	Location
Theoretical Structural Evaluation	Review of literature to establish consistency with relevant works. Explanation of internal consistency.	Chapter 2, Chapter 3 Chapter 3
Empirical Structural Evaluation	Description of example problems, how the framework is particularized to solve them and what evidence about the hypotheses is provided by solving the examples this way.	Chapter 4
Empirical Performance Evaluation	Discussion of how the particularization of the framework is easily extensible for these problems. Discussion of how other approaches to model validation are not appropriate for these problems.	Chapter 4 Chapter 2
Theoretical Performance Evaluation	Discussion of how the conceptual framework is sufficiently rich to span the problem domain. Discussion of how the framework applies to problems other than the examples. Discussion of how the framework supports behavioral model validation methodologies that are consistent with the requirements stated in Section 1.3.2.	Chapter 5

chapters. The following subsections are a summary of these results and the discussion entailed by theoretical performance evaluation.

5.1.2 Theoretical Structural Evaluation

Theoretical structural evaluation involves answering two main questions:

- Are the hypotheses consistent with the existing literature?
- Are the hypotheses consistent with themselves?

These questions must evaluate to the affirmative for the hypotheses to be accepted as appropriate. According to the evidence presented in this thesis, the hypotheses are consistent with the existing literature and with themselves. The following is an explanation of why this is so.

Consistency with Existing Literature

The hypotheses should build upon concepts and methods from the existing literature. This increases confidence in the hypotheses are compatible with accepted prior results. Consistency with the literature is desirable to the extent that the literature is appropriate for the stated problem domain. Inconsistencies are acceptable if one can show the literature to have limitations that are addressed by the proposed hypotheses.

Consistency with existing literature is addressed mainly in three parts of the thesis:

- Section 1.2: The meaning of model validation.
- Section 2.1: Fundamental capabilities and limitations of model validation.
- Section 2.2: Existing conceptual framework for model validation

In each case, the proposed hypotheses are consistent with the existing literature. The following is a summary of these findings.

As regards the meaning of model validation:

- The meaning of 'model validation' is stated in Section 1.2.2. Several key phrases in the definition are examined more closely and compared to several definitions for 'model validation' from the literature (given in Table 1.2. Specific phraseology varies among the definitions. For example, some authors use the phrase "domain of applicability" instead of "context" or "study objectives" instead of "user needs." Despite these variations, one can observe that the definition adopted in this thesis is equivalent in substance to those in Table 1.2. The motivation for stating a definition particular to this thesis is to highlight the concepts important for this thesis (such as context and inaccuracy) rather than to change the meaning of the term. This is common in the model validation literature, as evidenced by the variety of definitions cited in Table 1.2.
- The application of the definition for model validation is evident in the two-step assessment process of H2. To be considered valid, a model must have a context that is compatible with that of the problem for which it is used and be sufficiently accurate for user needs. This is consistent with the stated definition and, by extension, consistent with the literature.
- The first hypothesis also is consistent with this meaning of model validation. This is evidenced by the knowledge acquisition requirements stated in H1. It states that model creators must formalize their knowledge about the accuracy of a model over some well-defined context. This hypothesis includes no explicit restrictions on how one performs or interprets the model validation process. However, an assumption about what constitutes a valid model underlies the knowledge acquisition requirements. This hypothesis includes the elements of model validation according to the definition given in Section 1.2.2. Since that definition is consistent with the literature, so is H1.

As regards the fundamental capabilities and limitations of model validation:

- Section 2.1 is a discussion about the fundamental limitations of model validation. A main result of this section is that one cannot prove the validity of a model in a concrete mathematical sense. The reasoning for this is rooted in Hume's problem of induction. Essentially, one cannot "prove" a model to be valid based on a finite number of observations. Induction is the generalization of a finite number of observations to a general rule. However, there is no empirical basis for induction. One must make a subjective "leap of faith."
- A casual glance at this thesis might lead one to conclude that the proposed hypotheses are not compatible with the fundamentals of model validation on the grounds that they constitute an attempt to rigorously prove model validity. However, the hypotheses include no such pretense. Although the assessment steps of H2 can involve rigorous evaluations of mathematical statements, there are two reasons why this is not a contradiction. First, the hypotheses do not confine one to use such assessment methods. The examples of Chapter 4 involve such methods (e.g., see Section 4.2.2.), but they are not chosen because the hypotheses require them. Second, rigorous operations are not necessarily an indication of an attempt to prove validity. One performs such operations on formalizations of human knowledge—validity descriptions of a model and specifications for model applications. Subjectivity enters the picture during the formalization of this knowledge. Rigorous operations indicate an attempt to prove validity only in instances where one disregards the fallibility of human knowledge. Although the hypotheses do not include a statement denouncing attempts to prove validity, they include no statement in favor of it. Thus, responsibility rests with those involved in the model validation process and the hypotheses do not contradict the fundamentals.
- Although one cannot prove validity in a mathematical sense, model validation can be a scientific process. The basis for this is described in Section 2.1.2 and summarized in Section 2.1.3. Modern interpretations of science are that theories may be proven false, but may never be proven true. The roots of this perspective

are in Popper's notion of falsification. For model validation to be scientific, there must exist a means by which one can refute the validity of a model.

- The proposed hypotheses are consistent with a scientific process for model validation. For validity characterization of H1, one can establish a validity description through repeated attempts to refute a proposed description. For the assessment steps of H2, one can reject a model as valid for a particular use based on failure to contradict the statements "the model is context-compatible with the simulation problem" and "the model is sufficiently accurate for user needs."

As regards the existing conceptual framework for model validation:

- Section 2.2 is a review of the prevailing view on model validation. Section 2.2.2 is a description of the prevailing conceptual framework for model validation. A key result is that the conceptual framework is based on a particular conceptualization of a modeling and simulation process. This process is reflected in the depiction of the conceptual framework given in Figure 2.2 and leads to particular model validation sub-processes, such as verification and conceptual model validation.
- Section 2.3. is an examination of model development and validation from a reuse perspective. Section 2.3.1 includes a description of a modeling and simulation process is that corresponds to the model reuse scenarios of Section 1.3.1. Figure 2.3 is a depiction of this process.
- The process of Figure 2.3 is incompatible with the one underlying Figure 2.2. Therefore, model validation approaches based on the framework of Figure 2.2 are not generally appropriate for reuse scenarios. A more general interpretation of model validation is required.
- One can interpret the conceptual framework of Figure 2.2 in terms of the process depicted in Figure 2.3. Essentially, one can associate methods based on the framework of Figure 2.2 with model development. Subsequent reuse of a model

requires an extension to the framework. This is depicted in Figure 2.4. Figure 2.5 is an elaboration that includes concepts from the hypotheses of this thesis.

- Since the conceptual framework of Figure 2.2 can be inappropriate for model reuse scenarios, some incompatibility with the literature is necessary in order to answer the research questions of this thesis. However, outright rejection of the existing framework is unproductive. The conceptual framework proposed in this thesis is an extension of the framework of Figure 2.2 to situations not considered by its creators. Figure 2.5 serves as evidence of this. Thus, the proposed hypotheses address the limitations of the existing framework within a particular problem domain without contradicting it in other circumstances.

For the reasons stated above, one can be confident that the proposed hypotheses are consistent with the existing literature with respect to the meaning of model validation, the fundamental capabilities and limitations of model validation and existing conceptual framework for model validation.

Internal Consistency of Hypotheses

Internal consistency is necessary for the proposed hypotheses to stand as a combined body of work rather than individual statements. The hypotheses must incorporate the validation concepts in a way that is meaningful and appropriate. They must provide for a flow of validation-relevant knowledge that enables model users to perform model validation for a model they did not create. Sections 3.1 through 3.3 contain elaborations on the concepts involved in hypotheses. Section 3.4 contains a discussion of how these concepts fit into a model validation process and an argument for the internal consistency of the hypotheses. The following are highlights of the argument.

- Section 3.1 is an elaboration of the role of validity descriptions in the proposed conceptual framework. They are an unambiguous representation of validation-relevant knowledge from a model creator. They are comprised of a statement about the inaccuracy of a model in a specific context.

- Section 3.2 is an elaboration of context. It is interpreted as the limited set of situations in which some statement is true (Section 3.2.1). In model validation, one applies this concept when considering whether a candidate model truly represents the situation that one must model (Section 3.2.2).
- Section 3.3 is an elaboration of inaccuracy. It is interpreted as the total uncertainty present in a model (Section 3.3.1). In model validation, the inaccuracy of a model must be sufficiently small for the needs of model users (Section 3.3.2).
- Figure 3.4 is a flow chart for a model validation process based on the preceding concepts. It reflects the model reuse process depicted in Figure 2.3. Section 3.4.1 includes a description of the correspondence between the process steps and the hypotheses.
- As identified in Section 3.4.2, the process steps are consistent with one another. At the inception of each step, the requisite knowledge and information is available so that one can carry out the step. Assuming the individual steps are executed in a sound manner—e.g., creators produce well-formed validity descriptions, compatibility assessment is error free, etc.—the model validation process will proceed appropriately.

Based on the argument recounted above, one can be confident that the hypotheses and proposed conceptual framework are internally consistent.

5.1.3 Empirical Structural Evaluation

Empirical structural evaluation involves building confidence in the appropriateness of the example problems. In cases where proposed hypotheses reflect a specific method for solving a problem within some domain of similar problems, the primary concern for empirical structural evaluation is whether the examples reflect the characteristics of the problem domain. However, the topic of this thesis is a conceptual framework for a problem domain rather than one particular method. In this case, the examples serve to build confidence that the framework is appropriate for the problem domain. To do this, a particular method is adopted that is compatible with the proposed conceptual framework. Successful completion of the example problems using this method is a necessary step in building confidence in the framework. It serves as a needed "proof of concept."

For this thesis, empirical structural evaluation involves considering two main issues:

- Appropriateness of the examples with respect to the problem domain.
- Appropriateness of the adopted solution method with respect to the hypotheses.

These issues are addressed in Sections 4.1.1 and 4.1.2, respectively. The following is a summary of the results.

Appropriateness of the Examples

The appropriateness of the examples rests on how completely they, when taken together, reflect the characteristics of the behavioral model validation in reuse scenarios.

- The salient features of the problem domain are discussed at various points of the thesis and summarized in Section 4.1.1. The main criteria are that model creators have no specific knowledge of how someone might reuse their models, model users have no specific validation-relevant knowledge about a model beyond what is included in its validity description and model users may not have access to empirical data relative to which they can perform model validation.
- The example of Section 4.2 involves the development and use of a model. First, a model is developed and characterized. The result is a model and a corresponding validity description. Next, the model is used for a particular problem. The input

to this process is the model, its validity description and a description of the modeling problem. During the model development process steps (including validity characterization), no specific knowledge about any particular use of the model is considered. During the model use process steps (including compatibility and adequacy assessment), no knowledge about the model is considered beyond what is contained within its validity description. Furthermore, no empirical data is used.

- The example of Section 4.3 involves only the development and validity characterization of a model. The model in this example is more complex than that of the previous example. Although this example involves only the first of the problem domain characteristics (i.e., there is no knowledge about particular model uses during model development), it is a deeper treatment of validity characterization for a more complex model.
- Although the second example involves a more complex model than the first, both models are relatively simple relative to typical engineering problems. This limits the extent to which one can generalize the results of this thesis, but does not mean the examples are inappropriate. The objective underlying the example problems is to build confidence in the consistency and usefulness of the proposed conceptual framework. As such, the example problems involve simple models that are familiar to most engineers. This allows readers to focus on the framework concepts and their relationships rather than the details of a particular model.

Taken together, the example problems span the set of problem domain characteristics and serve as a basis for demonstrating the consistency and usefulness of the proposed conceptual framework. One can be confident in their appropriateness.

Appropriateness of the Adopted Method

To be appropriate for the example problems in this thesis, a model validation method must be based upon the proposed conceptual framework. It must incorporate the necessary concepts and reflect their relationships as defined in this thesis. The adopted method is described in Section 4.1.2 and its appropriateness with respect to the hypotheses is summarized below.

- The method includes a specific step in which one performs validity characterization. During this step, one defines a context and then determines the inaccuracy of the model within this context. This reflects the requirements of H1.
- Context is represented by bounds on variable values. This is a special case of the set-based representation used in the discussion of Section 3.2.
- Inaccuracy is represented by the magnitude of an uncertainty parameter. Magnitude is determined using a Euclidean norm. This representation is associated with epistemic uncertainty as described in Section 3.3.
- The method includes a compatibility assessment step (step 1 of H2) in which one compares the context of the model (as stated in its validity description) to the context of its use (as defined in the model use problem). This comparison is performed according to the semantics described in Section 3.2.2: a model is context-compatible with a use if the context of the use is a subset of that for the model.
- The method includes an adequacy assessment step (step 2 of H2) in which one compares the inaccuracy of the model (as stated in its validity description) to user needs (as defined in the model use problem). The magnitude of the uncertainty parameter must be less than the maximum tolerable inaccuracy for the user needs.
- A model is considered not valid for a particular use if it fails either compatibility assessment or adequacy assessment.

Given the above points, one can be confident that the adopted method is an appropriate implementation of the proposed conceptual framework and therefore compatible with the hypotheses.

5.1.4 Empirical Performance Evaluation

In this thesis, empirical performance evaluation involves establishing that the proposed conceptual framework is useful for solving the example problems. Support for this falls into two categories:

- Limitations of existing approaches in stated problem domain.
- Success of a method based on the proposed conceptual framework on the example problems.

These issues are addressed in several locations in the thesis. The following is a summary of the findings.

Limitations of Existing Approaches

The proposed hypotheses must overcome the limitations of existing approaches to model validation in order to have value. The limitations of existing approaches stems from their being based on a conceptual framework that is inappropriate for model reuse scenarios. The discussion of Section 5.1.2 includes a summary of these limitations. They are briefly reviewed below.

- The prevailing conceptual framework for model validation is depicted in Figure 2.2. In Section 2.2.2, it is established that this conceptual framework corresponds to a specific conceptualization of the modeling and simulation process. Model validation approaches based on this conceptual framework implicitly assume this process.
- Section 2.3 contains a description of a modeling and simulation process that corresponds to model reuse scenarios. This process is depicted in Figure 2.3. This process differs from the one underlying the framework of Figure 2.2.

- There are model reuse scenarios in which the framework of Figure 2.2 results in difficulties for model validation. The core problem is the separation of validation-relevant knowledge from those in the model validation process who require it. Several such problems arising in reuse scenarios are described anecdotally in Section 1.3.
- Figure 2.5 is an integration of the process of Figure 2.3 with the conceptual framework of Figure 2.2 and the conceptual framework proposed in this thesis. One can observe from this figure that the existing conceptual framework is not sufficient to handle reuse scenarios in general.

Thus, one can be confident that existing model validation methods—which are based on the framework of Figure 2.2—have limitations with regard to model reuse scenarios. Although these methods can work in some reuse scenarios, engineering designers require different methods for general reuse scenarios.

<u>Success on Example Problems</u>

To be judged a success, one must be able to perform the specified model validation tasks on the examples using a method based on the proposed conceptual framework. The criteria for success depends on the model validation task. For validity characterization, one must be able to formalize validation-relevant knowledge about a model. The resulting formalization—i.e., the validity description—must represent the known inaccuracy of the model over a well-defined context. For compatibility and adequacy assessment, one must be able to carry out the assessment comparisons using the validity description as the only source of knowledge about the model. The results are summarized below.

As regards validity characterization:

- In the example of Section 4.2, validity characterization is performed for a formulation of Newton's law of motion under the assumption that the rate of mass change is "negligible" (Section 4.2.1). The semantics of this assumption are conveyed in an unambiguous fashion by constructing a validity description of the model. This is done relative to a more general formulation of Newton's law that accounts for time-varying mass. The resulting validity description specifies the inaccuracy of the model over a well-defined context and conveys what the model creator means by "negligible".
- Section 4.2.3 includes a remark about the referent used during validity characterization. For the example of Section 4.2, there exists a relationship that is more general than the one used as a referent for the validity characterization reported in Section 4.2.1. Specifically, one could have used a relativistic formulation of the law of motion. That additional knowledge exists and is unused during validity characterization is not an indication of flaws in the proposed conceptual framework or the validity characterization process of Section 4.2.1. The key point is that given *some* knowledge and a model, one can perform validity characterization relative to *that* knowledge. Although important in its own right, the appropriateness of referent knowledge a separate issue.
- The example of Section 4.3 involves a model for extension of a beam held in axial tension. This model is more complex model than the one of Section 4.2. It involves several assumptions that impact its accuracy, such as the negligibility of thermal effects and that the extension of the beam is small. Given knowledge about the impact of these assumptions, one is able to formulate a validity description for the model that specifies its inaccuracy over a well-defined context.

As regards compatibility and adequacy assessment:

- The first example problem involves compatibility and adequacy assessment. Section 4.2.2 is a report of these steps. The example is to decide whether a particular model is valid for use in a given analysis problem. The stated analysis problem includes a context of interest and a desired prediction accuracy.

- Using the validity description for the model from Section 4.2.1 as the only knowledge about the model, one is able to determine that the context of the intended use is a subset of the context of the model. Thus, the model is judged to be context compatible with the intended use.
- Using the validity description for the model from Section 4.2.1 as the only knowledge about the model, one is able to determine that the model is sufficiently accurate for the stated user needs. Thus, the model is judged to be adequate for the intended use.
- Because it is both context compatible with and adequate for the intended use, the model is judged to be valid for the intended use. One reaches this conclusion using a validity description as the only source of knowledge about the model.

Given the above points, one can be confident that the model validation tasks were performed successfully using a method based upon the proposed conceptual framework.

5.1.5 Theoretical Performance Evaluation

Theoretical performance evaluation involves assessing the appropriateness and usefulness of the proposed conceptual framework to problems beyond those included in the examples. Thus, it is a generalization process that requires one to extrapolate beyond the results of Chapter 4. In this thesis, theoretical performance evaluation involves answering two main questions about the proposed conceptual framework:

- Is the conceptual framework appropriate for other simulation problems?
- Can the framework serve as a basis for methodologies consistent with the requirements of Section 1.3.2?

Here, "simulation problem" refers to the model and other circumstances involved in a reuse scenario. Variations from the example problems might include use of more complex models, different mathematical formalisms and availability of different knowledge during validity characterization (e.g., empirical data, reference models, etc.).

General Appropriateness of the Framework

The first of the above questions deals with the capabilities of the proposed conceptual framework from a conceptual and logical standpoint. At issue is whether the framework is sufficiently rich to span the problem domain. The characteristics of model validation for behavioral models in reuse scenarios are described in Section 4.1.1. The three main properties are:

- Model creators have no specific knowledge of how their models might be used.
- Model users have no specific validation-relevant knowledge about a model other than what is included in a validity description.
- Model users may not have access to empirical data relative to which they can perform model validation.

One can judge the proposed framework to be generally appropriate for the problem domain if one can conclude that it is appropriate for problems with each of the above characteristics. This is true of the proposed framework:

- Although they are simple relative to engineering problems, the examples are representative of the model validation problem in reuse scenarios. This point is made during the empirical structural evaluation of Section 5.1.3. The first example includes all three characteristics and the second example includes the first characteristic. Thus, the examples span the problem domain characteristics.

- A method based on the proposed conceptual framework is successful in solving the example problems. This point is made during the empirical performance evaluation of Section 5.1.4. Since this method solves a set of examples that span the problem domain, one can be confident in the appropriateness of the conceptual framework for the problem domain.
- Whether one can solve problems with different models and different types of reference knowledge is an issue of methodology rather than framework appropriateness. Introduction of different models or reference knowledge can require more sophisticated representations and methods, but the basic concepts and their relationships within the framework remain constant. As a demonstration of this, the example problems include validity characterization of two unrelated models in Section 4.2.1 and 4.3.3. Although the details differ in each case, both validity characterization efforts involve the same concepts and relationships.
- It is possible that other models require concepts and process steps *in addition* to those defined in the proposed hypotheses. One cannot rule out this possibility based on the examples and analysis of this thesis. However, this possibility does not undermine the appropriateness of the proposed conceptual framework. Essentially, appropriateness aligns with necessity rather than sufficiency, which is a stronger claim. To support such a claim about a conceptual framework, one requires a large body of evidence that is beyond the scope of any one thesis.

Thus, given the results of this thesis one can be confident that the conceptual framework is appropriate for general behavioral model reuse scenarios.

Methodology

The second question involved in theoretical performance evaluation deals with whether one can devise useful methodologies based on the proposed conceptual framework. This brings into question the availability of knowledge representations and methods that fit

Table 5.2: Requirements from Section 1.3.2.

No.	Requirement
1	The time spent performing validation activities at the point of model use must be made small.
2	Validation-relevant knowledge must be represented explicitly and associated with behavioral models.
3	Validation-relevant knowledge must be described in terms of concepts that have well-defined semantics that are independent of any particular person, group or project.
4	Validation-relevant knowledge must be expressed in a mathematically formal manner.

within the framework as well as whether a resulting methodology meets the requirements described in Section 1.3.2. These requirements stem from the potential challenges of behavioral model reuse scenarios in engineering design. They are repeated in Table 5.2.

As regards methods and knowledge representations:

- The appropriateness of particular representations and methods for implementing the framework depends on the properties of a model and simulation problem. This also is true for non-reuse scenarios and the framework of Figure 2.2. As a group, the methods associated in the literature with the framework of Figure 2.2 are appropriate for a wide range of models and simulation problems. To a limited extent, this supports the notion that one can find appropriate representations and methods for the framework proposed in this thesis. However, one can interpret the proposed framework as an augmentation of that of Figure 2.2. As such, one requires new or augmented methods to particularize it for use.

- This thesis lacks evidence to support claims about the general viability of methodologies based on the proposed framework. This a limitation of this work and is discussed further in Section 5.3.2. To support such a claim requires extensive evidence consisting of diverse example problems. Each example must include a different type of model or simulation problem along with an appropriate particularization of the framework. As discussed in Section 3.3, representations of inaccuracy remain a topic of research. The same is true of advanced methods for performing each of the three main tasks: validity characterization, compatibility assessment and adequacy assessment. Proper substantiation of general claims about methodologies requires evidence that is beyond the scope of any one thesis or dissertation.
- The examples in this thesis do provide evidence about a particular set of problems. They support the claim that one can easily particularize the framework to problems involving models expressed in closed form for which a higher-fidelity closed-form model is known. Furthermore, the examples support particularizations using interval-based context representations and inaccuracy representations involving an additive inaccuracy parameter with an interval bound. Although the examples do not span all possible engineering models and simulation problems, they cover a useful subset of the space and represent a point of departure toward future examples.

As regards Requirement 1 (Table 5.2):

- This requirement addresses the cost of performing model validation activities in the model use process. It specifies that the time spent performing such activities be small. This requirement is imprecise and therefore difficult to evaluate. Essentially, it means that all validation activities that are specific to the model (as opposed to a particular application of the model) should be performed outside of the model use process. From this perspective, one can consider the requirement to be met if within the proposed conceptual framework the only model validation activities to come during the use process are those that one cannot perform outside of the use process.

- By examining the flow diagram of Figure 3.4, one can see that the model validation activities one must perform in the use process are compatibility and adequacy assessment.
- Compatibility assessment requires one to have knowledge about the context of the model and the context of the simulation problem. This is established in Section 3.2 and demonstrated in Section 4.2.2. Adequacy assessment requires one to have knowledge about the inaccuracy of the model and the amount of inaccuracy users can tolerate. This is established in Section 3.3 and demonstrated in Section 4.2.2. Both compatibility and adequacy assessment involve knowledge about the particular model use and therefore must occur in the model use process.
- Compatibility and adequacy assessment are the *only* validation activities defined in the proposed conceptual framework that occur in the model use process. Since these activities cannot be performed outside of the model use process, one can consider the first requirement to be met. The actual time spent performing compatibility and adequacy assessment will vary depending on the model, simulation problem and chosen methods and knowledge representations. However, one can be assured that by using a methodology based on this conceptual framework that only a minimum of validation activities occur in the model use process.

As regards Requirement 2:

- According to this requirement, one must make validation-relevant knowledge explicit and associate it with the model. The motivation for this comes from the problems that can arise when knowledge about a model resides informally with model creators.
- This requirement is met by the use of validity descriptions. By definition, they are explicit representations of validation-relevant knowledge that are associated

with behavioral models. Thus, any methodology involving validity descriptions satisfies this requirement.

As regards Requirement 3:

- This requirement addresses how one should represent validation-relevant knowledge. It states that one must describe validation-relevant knowledge in terms of semantically well-defined concepts that are independent of any particular person, group or project. This is motivated by the possibility that model creators are unavailable to answer the questions of model users.
- In the proposed conceptual framework, one represents validation-relevant knowledge using a validity description. The components of a validity description are a context and inaccuracy pair. Any methodology based upon this framework includes these concepts.
- That one uses the concepts of context and inaccuracy does not guarantee Requirement 3 is met. Context and inaccuracy are abstract concepts. For model validation, one must define context and inaccuracy in terms of concepts with concrete interpretations in the problem domain. It is possible to select concrete concepts that are meaningful only to relatively few designers. One must avoid this.
- The examples of Chapter 4 result in validity descriptions that satisfy Requirement 3. Consider the context chosen during validity characterization of a model for Newton's 2^{nd} law of motion (Section 4.2). This context involves the physical concepts of velocity and rate of mass change. These are concrete physical concepts that any engineer will understand. Problems arise when context is defined in terms of highly specialized concepts such as "velocity of part A42" or the rate of change in the sums of the masses of the components of a particular system. These specialized concepts may be meaningful to some designers, but they can render the context representation useless to others.

- Thus, whether Requirement 3 is met depends on the model creators and likely model users sharing a common vocabulary about a problem domain. Although determining an appropriate vocabulary can be a challenge for some domains, it is possible in principle. The examples of Chapter 4 bear this out. In each example problem, the validity description is formalized in terms of well-defined concepts familiar to most engineering designers. It seems that some problem domains defy the precise definitions of concepts enjoyed in physics. However, much of behavioral modeling is rooted in physics. This is reason to believe that designers in particular industries or domains can agree on common vocabulary for describing their behavioral models.
- Given the above points, one can conclude that it is possible to develop model validation methodologies based on the proposed conceptual framework that satisfy Requirement 3.

As regards Requirement 4:

- As with the previous one, this requirement addresses how one should represent validation-relevant knowledge. It states that one should state it in a mathematically formal manner. The motivation is that validity descriptions should be unambiguous in order to be most effective. Whereas Requirement 3 addresses semantic ambiguity, this requirement addresses *syntactic* ambiguity. Also, mathematically formal representations are processed on a computer more easily.
- As is the case with Requirement 3, one can define methodologies that are based on the proposed conceptual framework but that violate Requirement 4.
- The most significant challenge to meeting this requirement is the formalization of assumptions that one might typically deal with informally. For example, one might embody into a model the assumption "thermal effects assumed negligible." Even if one understands the meaning of the concepts in the statement, the statement is ambiguous. How small is small enough to constitute "negligible"?

- Although it can be challenging, in many cases it is possible to formalize the impact of assumptions in a mathematically rigorous fashion. This is demonstrated in the examples of Chapter 4. For the model of axial extension in a beam (Section 4.3), the impact of assumptions about thermal effects are formalized by bounding the variables that influence thermal effects and determining the inaccuracy within this range. This results in a mathematically formal and unambiguous validity description for the model.
- Given the above points, one can conclude that it is possible to develop model validation methodologies based on the proposed conceptual framework that satisfy Requirement 4.

Summary of Theoretical Performance Evaluation

The proposed conceptual framework is appropriate for simulation problems other than those presented in this thesis and it can serve as a basis for methodologies consistent with the requirements from Section 1.3.2. The main caveat associated with the proposed framework is that one can arrive at methodologies based on it that are inappropriate for a given problem or that violate one or more of the stated requirements. However, this does not undermine the value of the framework. The standards for evaluation for this thesis are plausibility and proof-of-concept. Thus, given this caveat, one can be confident that the proposed conceptual framework is generally appropriate and useful for the validation of behavioral models in reuse scenarios.

5.2 Contributions and Implications

This section is a discussion of the main contributions of this thesis and their most significant implications. Section 5.2.1 is a description of the primary research contributions of this thesis. Section 5.2.2. is a discussion of several implications of this work from a broader perspective.

5.2.1 Primary Contributions

The first chapter of this thesis includes a vision for the reuse of behavioral models in engineering design. A principal motivation for behavioral model reuse is efficiency. Ideally, it can be faster and less expensive to reuse an existing model than it is to develop a new one. However, a significant challenge to widespread behavioral model reuse is the problem of model validation. There is no value in reusing a model when one cannot establish its validity. Despite the central importance of model validation, the existing literature is inappropriate for many model reuse scenarios. This limitation of the literature serves as the principal motivation for this thesis and leads to the primary research question originally stated in Chapter 1:

> ***Primary Research Question:*** *How can engineering designers perform behavioral model validation in a way that supports model reuse?*

This question forms a basis for the contributions of this thesis. The literature lacks a satisfactory answer, thus necessitating novel ideas.

<u>Validation-Relevant Knowledge</u>

This thesis contains several contributions related to *validation-relevant knowledge*. These contributions are based on the observation that a major challenge to performing behavioral model validation in reuse scenarios is the potential for knowledge relevant to model validation to be unavailable where it is needed. The identification of context and inaccuracy pairs as the fundamental elements of validation-relevant knowledge is a significant step towards answering the primary research question. Armed with these concepts, model creators can describe the properties of their models that are salient to model validation. Furthermore, by doing so in a mathematically formal way, they can represent their validation-relevant knowledge unambiguously. Although these concepts are implicit in the literature, they are not investigated in detail. Doing so becomes particularly important only when one considers behavioral model validation from the perspective of reuse scenarios.

Other contributions relating to validation-relevant knowledge include descriptions of the properties of context and inaccuracy and the relationships between them. Chapter 3 includes restrictions on the semantics of context that serve to guide future researchers and practitioners. One result is that logical propositions such as “mass is assumed negligible” and “length is assumed much larger than width” are meaningless absent well-defined semantics for the magnitudes “negligible” and “much larger.” The definition of such semantics is a necessary step in making the knowledge representation unambiguous to model users. Another result is about what constitutes compatibility between the context of a model and that of a use. Section 3.2.2 includes a rigorous definition for context compatibility that enables model users to eliminate from consideration models that do not correspond to their situation of interest. Although mathematical rigor may not always be important in practice, the rigorous definition conveys to practitioners the meaning of context compatibility in an unambiguous fashion.

Model Validation Process for Reuse Scenarios

The abstract *model validation process for reuse scenarios* is another contribution of this thesis. Figure 3.4 is a flow chart for this process. It includes a clean distinction between tasks that are the responsibility of model creators and those that are the responsibility of model users. Although ad hoc approaches to model validation can succeed at times, they are unreliable in general. The clear delineation in responsibilities alleviates any potential for confusion. By performing this process, model creators and model users are able to collaborate efficiently and effectively.

One novelty of the process is that validity descriptions serve as the interface between model creators and model users for the purposes of model validation. No additional communication is necessary. This frees the model use and model development processes from one another. Model users can select, validate and use models quickly and effectively while model creators develop and characterize models on their own timeline. Model users can perform their tasks without consulting model creators. This is a major advantage. In many of the reuse scenarios of Section 1.3.1, model creators may be unavailable for consultation at the time of use. Armed with the knowledge that models can have useful lives after their creators have moved on, companies may be more inclined to invest in modeling and simulation technologies. With longer lifetimes for models, companies can amortize the cost of developing them over many more uses. The validity description interface even makes it possible for companies to specialize in model development and validity characterization. Other companies can buy their models and use them as black boxes (thus preserving proprietary implementation details) while still being able to validate the use of the model through its corresponding validity description.

The use of validity descriptions in the process as an interface between model creators and model users is an improvement over text-based documentation. Although they both can serve as representations for the same knowledge, documentation and validity descriptions have different practical implications. Validity descriptions are mathematically formal. They are unambiguous and intended as a complete account of the validation-relevant knowledge about a model. In principle, documentation can convey the same formal knowledge as a validity description with the same degree of completeness. In practice, documentation often is neither very formal nor complete. Perhaps more significantly, being mathematically formal, one can relatively easily express validity descriptions in computer-interpretable form. This is necessary for automating computations with validity descriptions. Another drawback of documentation-based approaches is that they are labor-intensive during both model development and model use. Model creators must author comprehensive documentation that model users must then assimilate and apply. In contrast, use of validity descriptions as an interface is labor-intensive only during validity characterization. Model creators must still apply their insight and expertise to develop a complete and formal validity description. However, model users potentially can automate compatibility and adequacy assessment. Given a validity description and an appropriate description of an intended model use, these process steps require minimal expertise. This is particularly important in cases where model users have many candidate models to consider or where the validity descriptions are particularly complex.

Conceptual Framework as a Basis for a Methodology

As a whole, the proposed *conceptual framework* is a contribution because it forms the basis of a methodology for validating behavioral models in reuse scenarios. A methodology should (Arthur, et al. 1986, Nance, et al. 1988):

1. Organize and structure the tasks comprising the effort to achieve global objectives.
2. Include methods and techniques for accomplishing individual tasks (within the framework of global objectives).
3. Prescribe an order in which certain classes of decisions are made and the ways of making those decisions that lead to the desired objectives.

The framework defined in the hypotheses of this thesis satisfies the first and third items in this list. The flow diagram of Figure 3.4 defines the structure and flow of the tasks for validating behavioral models in reuse scenarios. The decision points in the process are compatibility and adequacy assessment. Based on the outcomes of those decisions, one can identify whether a model is valid for a particular use, thus achieving the objective of model validation.

The literature includes methodologies for performing model validation. However, these are based on the framework depicted in Figure 2.2 and therefore are not generally appropriate for reuse scenarios. The framework proposed in this thesis is appropriate for reuse and therefore suited for serving as the basis for a methodology. It serves as a necessary first step on the road to defining a comprehensive methodology. It includes the vocabulary one requires to describe and think about the problem of validating behavioral models in reuse scenarios, the tasks one must complete and the order in which to complete them. The framework lacks sufficient depth to define all relevant tasks and their orderings, but it captures the high-level features of the problem domain so that researchers can elaborate the framework in the future.

5.2.2 Broader Implications

If the proposed hypotheses become accepted within the engineering community they will have implications on other engineering activities. The following are some highlights.

More Powerful Modeling and Simulation Tools

The formal representation of model properties is a facilitating step in the development of more powerful modeling and simulation tools. Broadly speaking, to increase the "intelligence" of computer-based tools one must formalize the knowledge and information required to reach "intelligent" decisions. Formal computer-interpretable representations of validation-relevant knowledge is a first step towards such tools for modeling and simulation in engineering design.

Validity descriptions are a major piece of the puzzle for implementing useful model repositories. Current tools allow engineers to search a database of models based on configuration data such as file names, creation and modification dates and the identify of authors (Mocko, et al. 2004). However, this data provides no basis for deciding whether a model is useful in a particular situation. Such a database is useful for engineers to keep track of models they have created, but is not particularly conducive to model reuse. Enhancing model descriptions to include validation-relevant knowledge allows engineers to validate the models they retrieve for their particular uses. If one automates compatibility and adequacy assessment, engineers can search large databases for appropriate matches to their problems quickly and efficiently. This allows for broader searches and helps reduce the time engineers spend on validation activities during a design process.

Validity descriptions also make possible analysis tools for simulation results. Simulation runs often yield a great amount of data that is time consuming and tedious for engineers to analyze manually. One of the analyses that engineers perform is to see whether the simulation trajectory violated any assumptions of the model or input information. Validity descriptions capture the semantics of modeling assumptions. Advanced post-processing tools can compare the data traces to the validity description as a final validation step of a simulation study. Similar analysis tools might identify alternative models that are likely to be valid or might identify which information, if obtained, would most greatly reduce prediction inaccuracy. Such analyses currently are a manual endeavor.

Compositional Modeling

Compositional modeling involves the development of one model by assembling, or composing, other models. One motivation for compositional modeling is that it enables rapid development of complex models. For design, compositional modeling is important when evaluating the performance of a system. Typically, designers decompose system requirements into those for subsystems, develop the individual subsystems and compose them into a final system. Presumably, engineers develop models for the subsystems while designing them. One could compose the subsystem models to form a system model just as one would compose the subsystems themselves. Engineers might even use compositional modeling at the subsystem level, relying on libraries of models for physical phenomena to compose subsystem models.

For model composition to be viable, one must be able to establish the validity of the composed model efficiently. The proposed framework makes this easier by making validation-relevant knowledge formal and explicit. One can use the validity descriptions of component models when developing a validity description for a composed model. Although one requires additional knowledge, the validity descriptions of the component models are a significant contribution toward validating the composed model.

Compositional modeling is an important area of future research. Section 5.4 includes a summary of many research issues related to validating composed models.

Accounting for Inaccuracy during Decision Making

Use of validity descriptions can lead to increased rigor in engineering decision making. One of the challenges of accounting for uncertainty in decision making is the lack of a quantified appraisal of uncertainty. The inaccuracy stated in a validity description is exactly that. If model creators perform validity characterization, model users can account for inaccuracy in their decision making by using the resulting validity descriptions. This is particularly important for risk-sensitive industries, such as aerospace, construction and defense.

A New Approach for Modeling and Simulation Education

If the proposed conceptual framework is adopted, the way in which engineers learn modeling and simulation must change. The change in education is more substantial than simply introducing a few new methods. Under the present educational paradigm for modeling and simulation, engineering students learn how to make assumptions relying on their expertise. Instructors help students develop this expertise about such issues as when they can neglect friction or drag, when they can treat a body as a point mass and when they can neglect thermal effects or other phenomena. The precise semantics and implications of such assumptions remain qualitative and, often, implicit. In order to perform validity characterization, engineers must be able to quantify the impact of modeling assumptions in a way that is semantically precise. In addition to new analysis methods, this involves a new way of thinking about model development.

Engineering students also must learn a new perspective on the objectives of model development. Traditionally, model development focuses on the requirements of a specific simulation problem. For reuse, model creators must consider opportunities for future reuse of a model. This is analogous to the ongoing shift in engineering design toward product platforms and reusable product components. Modern designers often make design decisions with respect to potential future needs as well as immediate requirements. As is the case with engineered products, ease of development and potential for reuse typically are opposing objectives. Model creators must learn to balance the difficulties of developing and characterizing very general models that have high reuse potential with the reduced reuse potential of more specific models that are easier to develop and characterize.

New Business Opportunities

The adoption and further development of the proposed framework would enable new types of businesses in the engineering community. Because the framework supports the validation of reusable behavioral models, companies can specialize in producing behavioral models. Presently, there are consultants who will perform customized modeling and simulation related tasks on a contract-by-contract basis. Such services will always have a market and the consultants will earn a premium for their customized services. With efficient and effective model reuse, companies could develop a catalog of behavioral models that engineers could search and purchase from online. Engineers can validate them for their particular use according to the proposed framework. These models would be less customized, but would be available instantly and be less expensive than customized models.

5.3 Limitations

This section is a critical analysis of the limitations of this thesis. This includes separate discussions of the inherent limitations of the hypotheses and the limitations due to the scope of the thesis. Section 5.3.1 is an analysis of limitations inherent to the hypotheses. These limitations are unavoidable, but do not necessarily devalue the hypotheses. Section 5.3.2 is an analysis of limitations of the thesis itself. These limitations include claims that one cannot substantiate with evidence contained in this thesis and represent opportunities for future work. This discussion serves as a lead-in for the next section, which is a discussion of directions for future research.

5.3.1 Fundamental Limitations

The hypotheses proposed in this thesis have several inherent limitations. These limitations are due to the nature of the hypotheses and one generally cannot overcome them. However, these limitations do not represent shortcomings of the work. They are fundamental challenges that accompany any model validation approach to some degree. Thus, this subsection is less a criticism of this thesis and more a discussion of general caveats that are important to understand. These fall into three main categories, all of which relate to knowledge:

- One cannot have perfect knowledge about a behavioral model.
- One cannot always know what one does not know.
- One cannot know the truth of a statement made by someone else.

These issues are described in the following.

Imperfect Knowledge

One cannot have perfect knowledge about a behavioral model. This is a consequence of the problem with induction that Hume described. The review of epistemology and the philosophy of science in Section 2.1 includes a discussion of the problem of induction. Essentially, it is impossible to test exhaustively any realistic behavioral model. As such, the notion of proof of validity is fleeting.

This problem has a secondary consequence with respect to the proposed hypotheses. Just as one cannot prove the validity of a model, one cannot prove the correctness of a validity description. It is possible to disprove a validity description. All one requires is a counterexample—a point within the context where the inaccuracy is greater than stated. Proof of correctness of a validity description requires one to test every point within the context, which often is a continuous region.

One can think of this problem as a "second-order" problem. In this sense, the inaccuracy of a model is the first-order inaccuracy. The inaccuracy of the formalization of inaccuracy is a second-order inaccuracy. Conceptually, this is infinitely recursive. One can always speak about the $N+1^{th}$ order inaccuracy of the N^{th} inaccuracy statement. In practice, the significance of each subsequent level is orders-of-magnitude less than its predecessor. Thus, it is safe to stop the progression with a validity description (i.e., the first-order inaccuracy) given that the model creator ensures that the second-order inaccuracy is insignificant relative to that of the validity description.

This limitation is inherent to any model validation approach. Thus, it is not a drawback to the proposed hypotheses. In fact, the proposed hypotheses are superior to many other model validation approaches because they include the requirement that one formalize validity descriptions. By going through a validity characterization process, model creators are more likely to obtain an inaccuracy measure with lower second-order inaccuracy.

Unknown Unknowns

The problem of unknown unknowns is an important issue in model validation. Essentially, one often is unaware of deficiencies in one's own knowledge. It is a problem of not even knowing what question to ask.

Unknown unknowns is an insidious problem that is significant during validity characterization. Model creators typically are capable professionals that one can count on to express their knowledge faithfully and completely. This includes the things they know and also the things they know they do not know. The latter category reflects instances in which model creators are aware that an effect or phenomenon exists, but are unaware of its precise implications. In such cases, one can count on model creators to conduct the requisite research in order to answer their questions. However, problems arise when model creators are unaware that an effect or phenomenon exists. In such situations, they do not know what questions to ask.

The unknown unknowns problem plagues model validation efforts regardless of whether model reuse is an issue. However, it is a particularly challenging issue for model reuse. Model creators do not generally know the situations in which users will apply their models. They must do their best to explore the realm of feasible effects and interactions that can influence the accuracy of their models. No rigorous method exists for ensuring that one has identified all relevant factors. This is where creativity, resourcefulness and expertise are invaluable.

Trust

Whenever people collaborate, trust is a concern. Strictly speaking, one cannot know the truth of a statement made by another person without corroborating it. However, this is not practical in general. People must trust each other and each other's work products in order to be effective in a collaborative setting.

The issue of trust impacts model validation in both reuse and non-reuse scenarios. However, it is more significant in some scenarios because the mode of collaboration is less direct. In some scenarios, model users never meet or interact with the creators of their models. Since they have no personal knowledge of the trustworthiness of the model creators, model users must have some other basis for trusting their results.

The literature contains descriptions of two approaches for establishing a baseline of trust: accreditation and certification. The International Organization for Standardization (ISO) makes the following definitions (Rae, et al. 1995):

> Accreditation: A procedure by which a body of authority formally recognizes that a body or person is competent to carry out specific tasks

Certification: A procedure by which a third party gives written assurance that a product, process or service conforms to specified standards.

By this terminology, people and organizations can be accredited, while products, processes and services can be certified. Accreditation helps to identify companies and individuals that meet minimum standards on some task, such as modeling in a particular domain. Certification increases a user's confidence that a particular result—a validity description, for instance—is as specified. Alternately, certification can apply to the methods used to develop particular results. From a validation perspective, accreditation and certification can provide the basis for trust between creators and users.

Readers should note that this terminology is not strictly uniform. As Balci observes, the U.S. Department of Defense publishes a definition of accreditation that corresponds to the ISO definition for certification (Balci 2001). Although the transposition of terminology is unfortunate, it does not undermine the fundamental notions of providing assurances about the abilities of people and capabilities of methods

Accreditation and certification do not themselves imply trust. They simply are mechanisms by which designers can build confidence in their collaborators and/or the creations of their collaborators. Ultimately, it is up to the users of validity descriptions to decide whether they will trust them. Some designers may adhere to higher standards than others. In such cases, certification of a validity description may be a necessary, but insufficient requirement for a designer to use it. This is analogous to how some companies, such as military contractors and the aerospace industry, abide by tighter standards for safety and accountability than those is less safety-critical industries.

Individuals or organizations can be accredited to perform up to a standard set by industry-consensus or as established by an appropriate standards body, such as the ISO or the National Institute of Standards and Technology (NIST). Thus, any so-accredited designers could collaborate on model validation problems with minimal concern for the veracity of the validity descriptions shared among them. This provides a baseline level of trust that enables model users to apply deductive reasoning within the assessment steps. Instead of or in addition to the accreditation of people, individual validity descriptions or particular validity characterization methods can be certified by a similar standards body. Certification is particularly desirable for reusable behavioral models and their associated validity descriptions because it allows subsequent users to establish trust without necessarily knowing the identities of its creators. The basis for trust in such a situation rests with the model and validity description pair rather than with their creators.

5.3.2 Thesis Limitations

The following is a discussion of the scope and methodology limitations in this thesis and their implications. One can overcome such limitations by performing additional research. Thus, this discussion ties in closely with the discussion of future work in Section 5.4.

<u>Lack of Directly Useful Methods</u>

The scope of this thesis is limited to conceptual-level issues. The proposed hypotheses constitute a conceptual framework for behavioral model validation that is appropriate for reuse scenarios. The hypotheses do not include specific methods for performing model validation. This is a limitation of the practical usefulness of the hypotheses, it is consistent with the underlying objectives of the thesis. Although the thesis is of limited immediate use to practitioners, it serves as a roadmap for the future development of a practical methodology and can have a greater long-term impact than the development of a specific method.

Methods for model validation vary widely due to the breadth of models and modeling and simulation studies. This is evident when examining the traditional model validation literature and it applies to reuse scenarios. Given this breadth, the development of a particular method for model validation is useful only to a limited audience. In contrast, the development of a general framework for solving a class of problems can benefit a large number of people. Given the wealth of model validation literature for non-reuse scenarios and an understanding of the distinction between reuse and non-reuse scenarios, it is reasonable to attack the problem of developing a conceptual framework. This circumvents the growing pains that occur in a field when its members attempt to generalize a consistent framework out of a potpourri of ad hoc approaches.

Open Questions about Validity Descriptions

Although the validity description is a significant contribution in this thesis, there remain some open questions. Many of the questions relate to the practical issues associated with developing and computing with validity descriptions. For example, a comprehensive methodology for performing validity characterization would be a major contribution. As noted above, such issues are not a focus of this thesis and are left for future research.

The thesis also leaves open some fundamental questions. This is a reflection of the depths of the concepts involved in the framework. With regard to context, it is unclear from this thesis how well it scales to large and complex models. The number of quantities in a context may grow large for large or complex models such as those used in finite-element analysis and computational fluid dynamics. The models from the example problems are modest in size and complexity and therefore do not support any conclusions about this issue.

With regard to inaccuracy, it is unclear how one should represent and compare inaccuracies. This thesis includes descriptions of several alternative representations, but reaches no conclusion about which ones are appropriate. This is related to practical methodological issues, but potentially is a fundamental issue because the different representations have different semantics. The question about how to compare inaccuracies also is pertinent both practically and fundamentally. From a fundamental standpoint, some representations may simply be incompatible with others. This can be because they consist of incompatible mathematical formalisms (e.g., intervals versus probability) or because they have differing semantics (e.g., subjectivist probability versus frequentist probability).

The limited depth of this thesis reflects a focus on the overall framework. Proper investigations of context and inaccuracy might themselves constitute one or more theses. Section 5.4 includes a discussion of areas for future investigation related to validity description.

Limited Scope of Example Problems

The example problems of Chapter 4 are limited in scope. The models involved are of modest complexity and size. This is beneficial from the standpoint of illustrating the concepts of the framework and their relationships. However, it limits the extent to which one can generalize from the results. One such limitation is relates to the ease of particularization of the framework. Although the examples are consistent with the claim that one can easily particularize the framework to specific problems, there is insufficient evidence to claim that this generally is the case.

To overcome this limitation is not trivial. In order to tackle more complex examples, one requires methods appropriate for dealing with them. At a minimum, this involves adapting methods from the model validation literature. One then would have to establish that those methods are appropriate for use within the proposed framework. If one cannot identify and adapt methods from the literature, one must devise and evaluate novel ones. This is a significant undertaking and greatly expands the scope of the work. Essentially, it would add additional hypotheses to the thesis that address the issues of method and methodology. Involving two unknowns in the example problem (the framework and the method) can complicate hypothesis evaluation.

5.4 Directions for Future Study

One of the main contributions of this thesis is that it serves as a stepping stone to a great many other research questions relating to the validation of behavioral models in reuse scenarios. This section is a summary of several open questions worthy of investigation. The breadth of the topic precludes an exhaustive examination of potential research issues. The following is a selection of issues reflecting some of the broader implications from Section 5.2.2. Section 5.4.1 is a summary of research issues relating to representations for context and inaccuracy, methods for formalizing and computing with them and the development of comprehensive methodologies for validating behavioral models in reuse scenarios. Section 5.4.2 is a summary of research issues relating to the special topic of compositional modeling.

5.4.1 Representations, Methods and Methodology

This thesis contains basic descriptions of the fundamental concepts and relationships associated with performing behavioral model validation in reuse scenarios. With respect to the concepts, the focus is on describing their semantics rather than prescribing specific representations. Methods are not a focus of this thesis, being discussed only as required for the example problems. Methodologies are possible only once there exists a comprehensive body of methods. The following is a summary of research problems in these areas.

<u>Representations for Context and Inaccuracy</u>

To apply the conceptual framework, one must adopt specific representations for context and inaccuracy. These representations must be consistent with the semantic requirements described in Sections 3.2 (for context) and 3.3 (for inaccuracy). They also must correspond to computational methods that are tractable for the assessment steps (compatibility and adequacy) and for propagating them through to predictions.

The interval-based context representation from the example problems of Chapter 4 is a good candidate for general use, but requires further investigation. It is compatible with the semantic requirements of context, as is any set-based representation. An advantage of this representation is that compatibility assessment is computationally simple, involving only comparisons of corresponding variable bounds. Higher fidelity context representations can be a detriment if they require compatibility assessment algorithms that are computationally complex. A major research issue for context representation is the demonstration of the viability of this or other representations through advanced example problems. These should involve different types of models (e.g., static and dynamic, continuous-time and discrete-event, etc.) and different referents (e.g., empirical data, higher-fidelity models, expert knowledge, etc.).

Inaccuracy representation is a complicated issue that involves several open questions. One problem is to identify the different representations that are available and to describe their properties. Section 3.3 includes a brief discussion of different alternatives, but it does not constitute a comprehensive examination of the issue. For each representation, it is important to describe its representational power and to identify the situations for which it is appropriate. The existence of a single representation that is desirable for all problems is unlikely to exist. For instance, probability theory may be suitable when aleatory uncertainty dominates, but not when epistemic uncertainty is significant. This problem entails an extensive review and synthesis of the literature and analysis of the representations through theoretical or experimental means.

Another research problem associated with inaccuracy is to determine how engineers can deal with heterogeneous inaccuracy representations. As noted above, it is unlikely that one representation is suitable for all problems. This leads to the possibility that engineers must combine different inaccuracies that are stated using different representations. For example, one might have a model inaccuracy stated using a set-based formalism (e.g., the approach used in the example problems of Chapter 4) and model inputs with inaccuracies specified using probability theory. It is important to develop an understanding of which inaccuracy representations are compatible and how to incorporate them into the same problem.

Methods for Formalizing and Computing with Validity Descriptions

Research on specific methods for model validation is important. There exist significant research problems relating to the following tasks:

- Performing validity characterization
- Performing compatibility assessment
- Performing adequacy assessment

Validity characterization involves the formalization of validation-relevant knowledge about a model into a corresponding context and inaccuracy pair. This knowledge originates from an expert or from comparisons of a model to a higher-fidelity model or empirical data. Model creators may require different methods depending on the referent they use. For instance, statistical methods may be appropriate when they use empirical data, but not when they use experiential knowledge. Even for one type of referent, it is unlikely that one method is appropriate in all situations. It therefore also is important to develop understanding about the relative strengths of various methods and when each of them are best applied. Moreover, model creators require not just a suite of different methods, but a well-rounded methodology for validity characterization.

Compatibility assessment involves comparing knowledge about model context to knowledge about the model use context. The difficulty of doing this depends largely on the context representation. For a given representation, model users require efficient methods for determining whether one context subsumes another. The computational complexity of these methods should scale well with the number of variables constrained by the context.

Compatibility assessment may involve heterogeneous context representations. This can happen unless all engineers (both creators and users) agree a priori to use the same representation. In the event of heterogeneous representations, model users require methods for converting between representation types. This may not be possible in all cases and may require approximations in lieu of mathematically rigorous mappings.

Adequacy assessment involves comparing knowledge about model inaccuracy to the inaccuracy requirements of a problem. Research issues associated with adequacy assessment are similar to those associated with compatibility assessment. A particularly important problem for adequacy assessment is dealing with heterogeneous representations. The presence of different types of uncertainty (i.e., aleatory versus epistemic) in a problem can mean that different inaccuracies (e.g., that of the model, input variables, intended model sue, etc.) are formalized using different representations.

Adequacy assessment depends on user accuracy needs. Sometimes, users cannot determine precise accuracy needs until after selecting a model and performing simulation. For example, consider a situation in which users compare two alternatives, each of which they evaluate by performing a simulation. The level of model inaccuracy that is tolerable depends in part on the performance difference between the alternatives–something that is unknown until after performing simulation. An important research problem is the development of methods to estimate accuracy needs prior to performing simulation. Such methods should focus on ruling out obviously bad models. Although this would not allow users to identify a adequate models with certainty, it does allow them to focus on those that are more likely to be adequate.

<u>Methodologies for Validating Behavioral Models in Reuse Scenarios</u>

One contribution of this thesis is that it forms the foundation of a methodology for performing behavioral model validation in reuse scenarios. The proposed framework includes two of the three elements of a methodology. The final element—methods and techniques for accomplishing individual tasks—will result from sustained research in the area.

One can develop methodologies for validity characterization and the assessment steps independently. For validity characterization, a methodology must address questions about how to select representations and how to map validation-relevant knowledge into the selected representation. For the assessment steps, a methodology must address how to select appropriate methods and how to estimate unknown model use parameters. The selected representation impacts which methods are available to users. As such, a validity characterization methodology must address the potential tradeoffs that can exist between the desire for higher-fidelity representations and computationally efficient methods.

5.4.2 Compositional Modeling

Compositional modeling is an important behavioral model reuse scenario. However, the research community must address several open questions in order to improve the efficiency of compositional modeling. In relation to this thesis, the main problem is establishing the validity of a composed model.

A composed model consists of several interconnected component models. The validity description of a composed model depends on those of its component models. Assuming that the composed model is complete—that is, that all relevant phenomena are included—one can, in principle, derive its validity description from those of the components. For example, the context of a composed model is the smallest context common to all of its components (i.e., the intersection of the contexts). When they are available, other sources of validation-relevant knowledge—such as empirical data, existing models or domain expertise—also influence the validity description of a composed model. Researchers must develop methods for determining validity descriptions for composed models. Basic methods should yield validity descriptions based on those from component models. More advanced methods should handle situations in which other referents are available. Dealing with heterogeneous validity description representations is another important issue. One is likely to encounter different representations when composing several models.

One problem that researchers must address is determining whether a composed model is complete. This is a fundamental issue that can undermine validation efforts. Although one can develop a validity description for a composed model from those of its constituent models, this description may not be correct. This can happen when one neglects to include a model. The validity descriptions of the included models are not sufficient for one to account for the inaccuracy of a neglected model. Essentially, this is an unknown-unknowns problem: one cannot account for a phenomenon when unaware of it. Researchers must develop methods and methodologies for dealing with this issue.

5.5 Closing Remarks

Behavioral model validation is challenging endeavor. Its underlying principles follow from basic understandings of science and knowledge. According to these principles, it is impossible to prove the validity of a model. Despite this limitation, model validation can yield useful conclusions and is an essential part of any simulation study.

This thesis is an exploration of behavioral model validation for scenarios in which model users apply previously existing models to new situations. In these scenarios, engineers face challenges beyond those of ordinary model validation scenarios. One such challenge is the potential for engineers to lack the knowledge necessary to perform model validation. Model validation involves knowledge about the model and its use. In reuse scenarios, model creators determine the assumptions and limitations of a model, but cannot know the characteristics of all possible uses. In contrast, model users know how they will use a model, but lack knowledge about its underlying assumptions and limitations. Neither group alone is able to establish model validity.

This thesis contains the description of a conceptual framework for performing model validation in reuse scenarios. The framework includes the concepts and relationships engineers require to conceptualize the problem of validating behavioral models in reuse scenarios. The framework also constitutes an abstract process for performing model validation. Model creators formalize their validation-relevant knowledge about a model in the form of a validity description that is associated with that model. A validity description contains a statement quantifying the total uncertainty, or inaccuracy, of a model over a well-defined set of circumstances, or context. Model users compare the validity description of a model to the properties of their intended use of the model. Users consider a model to be valid for the intended use only if its context is compatible with the intended use circumstances and its inaccuracy is adequate for user needs.

This thesis is conceptual in nature. It serves more as a guide to the problem domain of validating behavioral models in reuse scenarios than as a guide to solving a specific problem within the domain. This is inline with thesis objectives. It represents a value judgment in favor of breadth over depth for this particular work. A major result of this thesis is that the conceptual framework underlying ordinary model validation methods is inappropriate for model reuse scenarios. This serves as motivation for a broad synthesis of ideas.

Conceptual frameworks are fundamental to any problem domain. When developing methods for a particular problem, researchers have in mind—either implicitly or explicitly—a conceptual framework for the problem domain. Such a framework is necessary because it provides meaning to the method. Thus, the value of this thesis: it serves as a link between the concrete problem domain and the abstract mathematics in which one defines specific methods.

In some instances, conceptual frameworks emerge from a body of individual works. This is what happened for ordinary model validation. Through a synthesis of much prior research, the Society for Computer Simulation Technical Committee on Model Credibility described a consensus view of an appropriate framework for model validation (Schlesinger, et al. 1979). With minor variations, this framework survives today and forms a basis for research on model validation.

Conceptual frameworks need not result from a generalization process. One can adopt a more proactive approach. In the case of validating models in reuse scenarios, it is possible to draw upon understandings of model reuse and ordinary model validation to describe an appropriate conceptual framework. Such is the approach of this thesis. An advantage of this approach is that it can have a greater impact on subsequent research. Assuming the framework is correct, researchers can use it as a guide when developing corresponding methods and methodologies. Even if the framework is incorrect in some way, it can serve as a seed for discussion and object for refinement in future work. Initial development of a unified view of the problem domain helps promote communication among researchers and help prevent unnecessary fragmentation within the community. The community can engage in a proactive discussion about how to define the problem domain rather than let it be defined by a handful of narrow successes.

One drawback of the approach adopted in this thesis is that support for the proposed framework is limited. The relatively modest scope of the example problems limits the generality of the conclusions one can draw. However, this does not undermine the contributions of this thesis. The standard for this thesis is plausible appropriateness, not absolute proof. The support for the proposed framework is sufficient to warrant future investigation. Although there always exists the possibility that the framework is inappropriate for some specific situations, it certainly is appropriate for some.

Since it contains no directly useful methods, the ultimate value of this thesis rests with the future work that is based upon it. The more valuable are these works, the more valuable is this thesis. However, it presently is impossible to value the thesis in this way. One only can do so retrospectively, after the subsequent works are complete. The best one can do now is consider the potential of future work that is enabled by the proposals in this thesis. Section 5.2.2 is a discussion of several broader implications of the proposed framework. These include more powerful modeling and simulation tools, widespread compositional modeling, accounting explicitly for inaccuracy in decision making and potential new business opportunities for model-creating companies. This is in addition to the basic implication that engineering designers can reuse behavioral models quickly and effectively, thereby improving engineering design process. Although not directly a solution to any of them, the proposed conceptual framework serves as a roadmap for thinking about and solving these problems.

REFERENCES

AIAA. Guide for the Verification and Validation of Computational Fluid Dynamics Simulations. AIAA-G-077-1998. American Institute of Aeronautics and Astronautics, Reston, VA. 1998.

Akman, V. and M. Surav. "Steps Toward Formalizing Context." AI Magazine 17.3 (1996): 55-72.

---. "The Use of Situation Theory in Context Modeling." Computational Intelligence 13.3 (1997): 427-38.

Antonsson, E. K. and K. N. Otto. "Imprecision in Engineering Design." ASME Journal of Mechanical Design 117 (1995): 25-32.

Arthur, J. D., R. E. Nance and S. E. Henry. A Procedural Approach to Evaluating Software Development Methodologies: The Foundation. Tech. Rep. TR-86-24. Virginia Polytechnic Institute and State University, Blacksburg, VA. 1986.

Balci, O. "Credibility assessment of simulation results: The state of the art." Conference on Methodology and Validation, Orlando, FL. Ed. O. Balci. Society of Computer Simulations, 1987. 19-25.

---. "Validation, Verification and Testing Techniques throughout the Life Cycle of a Simulation Study." Annals of Operations Research 53 (1994): 121-73.

---. "Principles and Techniques of Simulation Validation, Verification and Testing." Winter Simulation Conference, 3-6 December 1995. Eds. C. Alexopoulos, W. R. Lilegdon and D. Goldsman. Piscataway, NJ: Institute of Electrical and Electronics Engineers, 1995. 147-54.

---. "Principles of Simulation Model Validation, Verification and Testing." Transactions of the Society for Computer Simulation International 14.1 (1997): 3-12.

---. "Verification, Validation and Accreditation." Winter Simulation Conference. Eds. D. J. Medeiros, E. F. Watson, J. S. Carson and M. S. Manivannan. Piscataway, NJ: Institute of Electrical and Electronics Engineers, 1998. 41-8.

---. "A Methodology for Certification of Modeling and Simulation Applications." ACM Transactions on Modeling and Computer Simulation 11.4 (2001): 352-77.

Banks, J., D. Gerstein and S. P. Searles. "Modeling Processes, Validation and Verification of Complex Systems: A Survey." Conference on Methodology and

Validation, Orlando, FL. Ed. O. Balci. Society for Computer Simulation, 1987. 13-8.

Barlas, Y. and S. Carpenter. "Philosophical Roots of Model Validation: Two Paradigms." System Dynamics Review 6.2 (1990): 148-66.

Ben-Haim, Y. Information Gap Decision Theory. Series on Decision and Risk, Ed. Y. Ben-Haim. London: Academic Press, 2001.

Birta, L. G. and F. N. Ozmizrak. "A Knowledge-based Approach for the Validation of Simulation Models: The Foundation." ACM Transactions on Modeling and Computer Simulation 6.1 (1996): 76-98.

Caughlin, D. "Verification, Validation and Accreditation (VV&A) of Models and Simulations through Reduced Order Metamodels." 1995 Winter Simulation Conference. Eds. C. Alexopoulos, K. Kang, W. R. Lilegdon and D. Goldsman. Piscataway, NJ: Institute of Electrical and Electronics Engineers, 1995. 1405-11.

Chandrashekar, N. and S. Krishnamurty. "Bayesian Evaluation of Engineering Models." ASME Design Automation Conference, Montreal, Canada, 29 Sept. - 2 Oct. DETC2002/DAC-34141. 2002.

Christen, E. and K. Bakalar. "VHDL-AMS - A Hardware Description Language for Analog and Mixed-signal Applications." IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing 40.10 (1999): 1263-72.

Clayton, M. J., P. Teicholz, M. Fischer and J. Kunz. "Virtual Components Consisting of Form, Function and Behavior." Automation in Construction 8.3 (1999): 351-67.

Dubois, D. Possibility Theory. Trans. E. F. Harding. New York: Plenum Press, 1988.

Falkenhainer, B. and K. D. Forbus. "Compositional Modeling: Finding the Right Model for the Job." Artificial Intelligence 51 (1991): 95-143.

Ferson, S. "Probability Bounds Analysis Solves the Problem of Incomplete Specification in Probabilistic Risk and Safety Assessments." Proceedings of the Ninth Conference on Risk-Based Decision Making in Water Resources IX, Santa Barbara, CA, United States, 15-20 October. Eds. Y. Y. Haimes, D. A. Moser, E. Z. Stakhiv, G. Zisk and B. Zisk. American Society of Civil Engineers, 2000. 173-88.

Ferson, S. and L. Ginzburg. "Different Methods are Needed to Propagate Ignorance and Variability." Reliability Engineering and System Safety 54.2-3 (1996)

Gero, J. "Design Prototypes: A Knowledge Representation Schema for Design." AI Magazine 11.4 (1990): 26-36.

Guha, R. V. and D. B. Lenat. "Language, Representation and Contexts." Journal of Information Processing 15.3 (1992): 340-9.

Guha, R. V. and J. McCarthy. "Varieties of Contexts." Modeling and Using Context, 4th International and Interdisciplinary Conference, CONTEXT 2003, Stanford, CA, 23-25 June. Eds. P. Blackburn, C. Ghidini, R. M. Turner and F. Giunchiglia. Springer, 2003. 164-77.

Halliday, D. and R. Resnick. Fundamentals of Physics. 3rd ed. New York: John Wylie & Sons, 1988.

Hanson, K. M. and F. M. Hemez. "A Framework for Assessing Confidence in Computational Predictions." Experimental Techniques 25 (2001): 50-5.

Hazelrigg, G. A. "On the Role and Use of Mathematical Models in Engineering Design." ASME Journal of Mechanical Design 121.3 (1999): 336-41.

Herskovitz, P. J. "A Theoretical Framework for Simulation Validation: Popper's Falsificationism." International Journal of Modeling and Simulation 11.2 (1991): 56-8.

Hume, D. A Treatise of Human Nature. 1739-40. Oxford: Clarendon Press, 1965.

Kleindorfer, G. B. and R. Ganeshan. "The Philosophy of Science and Validation in Simulation." Winter Simulation Conference. Eds. G. W. Evans, M. Mollaghasemi, E. C. Russell and W. E. Biles. Piscataway, NJ: Institute of Electrical and Electronic Engineers, 1993. 50-7.

Knepell, P. L. and D. C. Arangno. Simulation Validation: A Confidence Assessment Methodology. Los Alamitos, CA: IEEE Computer Society Press, 1993.

Kuhn, T. S. The Structure of Scientific Revolutions. 1962. 3rd ed. Chicago, Il: University of Chicago Press, 1996.

Law, A. M. and W. D. Kelton. Simulation Modeling and Analysis. 3rd ed. Boston: McGraw-Hill, 2000.

Mattsson, S. E., H. Elmqvist and M. Otter. "Physical System Modeling with Modelica." Control Engineering Practice 6 (1998): 501-10.

McCarthy, J. "Notes on Formalizing Context." IJCAI-93: Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence, Chambery, France, 28 Aug.-3 Sept. Morgan Kaufmann Publishers, 1993. 555-60.

Meckesheimer, M., R. R. Barton, T. Simpson and A. J. Booker. "Computationally Inexpensive Metamodel Assessment Strategies." ASME Design Automation Conference, Pittsburgh, PA. DETC-2001/DAC-21028. 2001.

Miller, D. Critical Rationalism: A Restatement and Defence. Chicago, Il: Open Court, 1994.

Minsky, M. "Models, Minds, Machines." IFIP Congress. 1965. 45-9.

Mocko, G., R. J. Malak Jr., C. J. J. Paredis and R. Peak. "A Knowledge Repository for Behavioral Models in Engineering Design." ASME Computers and Information in Engineering Conference, Salt Lake City, UT. CIE2004-144. 2004.

Nance, R. E. "Model Development Revisited." Winter Simulation Conference. Eds. S. Sheppard, U. W. Pooch and C. D. Pegden. Piscataway, NJ: Institute of Electrical and Electronic Engineers, 1984. 75-80.

Nance, R. E. and J. D. Arthur. "The methodology roles in the realization of a model development environment." 20th Conference on Winter simulation, San Diego, California, United States, 12-14 Dec. Eds. M. A. Abrams, P. Haigh, L. and J. C. Comfort. Piscataway, NJ: Institute of Electrical and Electronics Engineers, 1988. 220-5.

Naylor, T. H. and J. M. Finger. "Verification of Computer Simulation Models." Management Science 14.2 (1967): B92-B101.

Oberkampf, W. L., S. M. DeLand, B. M. Rutherford, K. V. Diegert and K. F. Alvin. "Error and Uncertainty in Modeling and Simulation." Reliability Engineering and System Safety 75.3 (2002): 333-57.

Oberkampf, W. L. and T. G. Trucano. "Verification and Validation in Computational Fluid Dynamics." Progress in Aerospace Sciences 38 (2002): 209-72.

Ohanian, H. C. Modern Physics. 2nd ed. Englewood Cliffs, NJ: Prentice Hall, 1995.

Page, E. H., B. S. Canova and J. A. Tufarolo. "A Case Study of Verification, Validation and Accreditation for Advanced Distributed Simulation." ACM Transactions on Modeling and Computer Simulation 7.3 (1997): 393-424.

Pahl, G. and W. Beitz. Engineering Design: A Systematic Approach. 2nd ed. Trans. K. Wallace, L. Blessing and F. Baurt. Ed. K. Wallace. London: Springer-Verlag, 1996.

Parry, G. W. "The Characterization of Uncertainty in Probabilistic Risk Assessment of Complex Systems." Reliability Engineering and System Safety 54.2-3 (1996): 119-26.

Pedersen, K., J. Emblemsvag, R. Bailey, J. K. Allen and F. Mistree. "Validating Design Methods and Research: The Validation Square." ASME Design Theory and Methodology Conference, Baltimore, MD, 10-14 Sept. DETC2000/DRM-14579. 2000.

Popper, K. R. The Logic of Scientific Discovery. 1934. 6th ed. London: Hutchinson, 1972.

---. Realism and The Aim of Science. Totowa, NJ: Rowman and Littlefield, 1983.

Rae, A., P. Robert and H.-L. Hausen. Software Evaluation and Certification: Principles, Practice and Legal Liability. London, UK: McGraw-Hill, 1995.

Robinson, S. "Simulation Model Verification and Validation: Increasing The User's Confidence." Winter Simulation Conference. Eds. S. Andradottir, K. J. Healy, D. H. Withers and B. L. Nelson. Piscataway, NJ: 1997. 53-9.

Rosenberg, A. Philosophy of Science: A Contemporary Introduction. Routledge Contemporary Introductions to Philosophy, Ed. P. K. Moser. New York, NY: Routledge, 2000.

Sargent, R. G. "An Expository on Verification and Validation of Simulation Models." Winter Simulation Conference, San Francisco, CA. Eds. D. Gantz, G. Blais and S. Solomon. Piscataway, NJ: Institute of Electrical and Electronics Engineers, 1985. 15-22.

---. "An Overview of Verification and Validation of Simulation Models." Winter Simulation Conference, Atlanta, GA. Eds. A. Thesen, H. Grant and W. D. Kelton. Piscataway, NJ: Institute of Electrical and Electronics Engineers, 1987. 33-9.

Schafer, G. A Mathematical Theory of Evidence. Princeton, NJ: Princeton University Press, 1976.

Schlesinger, S., R. E. Crosbie, G. S. Innis, C. S. Lalwani, J. Loch, R. J. Sylvester, R. O. Wright, N. Kheir and D. Bartos. "Terminology for Model Credibility." Simulation 32.3 (1979): 103-4.

Schruben, L. W. "Establishing the Credibility of Simulations." Simulation 34.3 (1980): 101-5.

Shooter, S. B., W. Keirouz, S. Szykman and S. J. Fenves. "A Model for the Flow of Design Information in Product Development." Engineering with Computers 16 (2000): 178-94.

Simon, H. A. The Sciences of the Artificial. 3rd ed. Cambridge, MA: MIT Press, 1996.

Szykman, S., R. Sriram, C. Bochenek and J. Racz. "The NIST Design Repository Project." Advances in Soft Computing - Engineering Design and Manufacturing. London: Springer-Verlag, 1998.

Szykman, S., R. Sriram, C. Bochenek, J. Racz and J. Senfaute. "Design Repositories: Engineering Design's New Knowledge Base." IEEE Intelligent Systems 15.3 (2000): 48-55.

Szykman, S., R. Sriram and W. C. Regli. "The Role of Knowledge in Next-generation Product Development Systems." ASME Journal of Computing and Information Science in Engineering 1.1 (2001): 3-11.

Tiller, M. M. Introduction to Physical Modeling with Modelica. Kluwer International Series in Engineering and Computer Science. Boston, MA: Kluwer, 2001.

Tucker, W. T. and S. Ferson. Probability Bounds Analysis in Environmental Risk Analysis. Applied Biomathematics, Setauket, NY. 2003. <http://www.ramas.com/pbawhite.pdf >.

U. S. General Accounting Office. Guidelines for Model Evaluation. PAD-79-17. Washington, D.C.: GPO, January 1979.

U. S. Department of Defense. DoD Modeling and Simulation (M&S) Verification, Validation and Accreditation (VV&A). DoD Instruction Number 5000.61. Washington, D.C.: GPO, 13 May 2003. 24 Feb 2005 <http://www.dtic.mil/whs/directives/corres/pdf/i500061_051303/i500061p.pdf >.

"Valid" Oxford English Dictionary. 2nd ed. 1989.

Yager, R. R., J. Kacprzyk and M. Fedrizzi, Ed.^Eds. Advances in the Depster-Shafer Theory of Evidence. New York, John Wiley and Sons, 1994.

Zadeh, L. A. "Fuzzy Sets." Information and Control 8 (1965): 338-53.

Zeigler, B. P., H. Praehofer and T. G. Kim. Theory of Modeling and Simulation. 2nd ed. San Diego, CA: Academic Press, 2000.